BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
Please do not report the images to the
Image Problem Mailbox.

(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 054 391 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
22.11.2000 Bulletin 2000/47

(51) Int. Cl.[7]: **G10L 15/26**

(21) Application number: **00304267.8**

(22) Date of filing: **19.05.2000**

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **21.05.1999 GB 9911971**

(71) Applicant: **CANON KABUSHIKI KAISHA**
Tokyo (JP)

(72) Inventor: **Keiller, Robert Alexander, Canon R. C. Europe Ltd.**
Guildford, Surrey GU2 5YF(GB)

(74) Representative:
Beresford, Keith Denis Lewis et al
BERESFORD & Co.
High Holborn
2-5 Warwick Court
London WC1R 5DJ (GB)

(54) A system capable of processing speech data, a server for such a system and a machine for use in such a system

(57) A system has a first machine (9 to 11,12,13, 15) couplable to a network (N) and capable of carrying out at least one function, a speech data receiver (26, 27) for receiving speech data representing instructions spoken by a user and specifying a function to be carried out by the first machine and a speech data transmitter (27,28) for transmitting the speech data to a speech server (2) couplable to the network (N). The speech server (2) has a speech manager (6) for accessing a speech recognition engine (5) for performing speech recognition on speech data received over the network to produce recognised speech data, an interpreter (7) for processing recognised speech data to derive from the speech data commands for causing the first machine to carry out the function specified in the spoken instructions and a command transmitter (6) for transmitting said commands over the network to the first machine, the first machine having a control command receiver (27,28) for receiving control commands over the network (N) from the speech server (2) and a controller (20,27) responsive to the control commands for causing the function specified by the spoken instructions to be carried out.

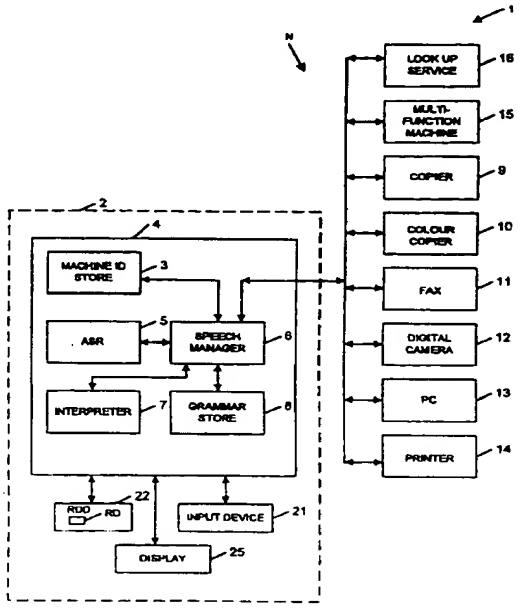


FIG. 1

EP 1 054 391 A2

Description

[0001] This invention relates to a system in which a number of machines are coupled together, in particular, but not exclusively, this invention relates to an office network system in which items of office equipment such as photocopi rs, facsimile machines, printers, personal computers and the like are coupled to a server.

[0002] In conventional systems such as office quipment n twork systems, instructions for controlling the operation of a machine connected to the system are either input directly to th machine itself using a control panel of the machine or are supplied via, for example, a personal computer also connected to the network.

[0003] The use of automatic speech recognition engin s in relation to computer programs or software such as word processing packages and the like is becoming more common. However, the use of such automatic speech recognition engines is generally in the form of specific automatic speech recognition engines at each personal computer on the network trained to the voice of the user of that personal computer.

[0004] It is an aim of the present invention to provide a system, a server for use in a system and a machine for use in such a system wherein a user can control the operation or functioning of a machine connected to the network, for example a copier or facsimile machine in the case of an office equipment network, using spoken commands.

[0005] In one aspect, the present invention provides a system having a server provided with speech recognition means and at least one machine couplable to a network, means for receiving spoken commands from a user, means for transmitting speech data representing the spoken commands to the server wherein the at least one machine has means for receiving from the server commands for controlling operation of that machine.

[0006] In an embodiment, the speech server has access to at least one grammar defining rules relating to the type of commands that can be input to a machine or network and words that can be used within those commands. In an embodiment, different grammars are provided for different types of machines so that, for example, in an office environment the server has access to a copy grammar for use with copying machines, a facsimile grammar for use with facsimile machines and a print grammar for use with printers.

[0007] In an embodiment, the server also has access to at least one shared grammar common to the grammars for the different types of machines.

[0008] In an embodiment, the server has access to a store of voice macros or predefined commands which associate a mnemonic or phrase with a set or series of functions to be carried out by a machine.

[0009] In an embodiment, at least one of the server and the machines couplable to the network has access to a look-up service storing information relating to the capability of machines on the network so that, for example, when a user requests from a machine a function that is not available at that machine the user can be provided with information identifying any machines on the network that can carry out the requested function.

[0010] In an embodiment, the server has access to a plurality of different automatic speech recognition engines and is arranged to select the most likely interpretation of received speech data on the basis of the results received from each of the automatic speech recognition engines.

[0011] In an embodiment, the speech server has means for accessing an automatic speech recognition ngine trained to the voice of the person from whom the spoken commands are received.

[0012] In an embodiment, speech data representing commands spoken by a user are supplied to the server via a telephone system associated with the network. Preferably, the telephone system comprises a DECT (Digital Enhanced Cordless Telecommunications) system.

[0013] Embodiments of the present invention will now be described, by way of example, with reference to the accompanying drawings, in which:

Figure 1 shows a schematic block diagram of a first embodiment of a system in accordance with the present invention;

Figures 2 to 6 are block diagrams showing, respectively, a copier, a facsimile machine, a digital camera, a personal computer and a printer for connection to the network shown in Figure 1;

Figures 7 to 11 show lists of words included in different grammars stored in a grammar store shown in Figure 1;

Figure 12 shows a flowchart for illustrating installation of a machine onto the network shown in Figure 1;

Figure 13 shows an example of data stored by a look-up service of the system shown in Figure 1;

Figure 14 shows a top level flowchart for illustrating functioning of a machine coupled to the network shown in Figure 1;

Figure 15 shows a top level flowchart illustrating functions carried out by a speech server of the system shown in Figure 1;

Figure 16 shows a flowchart for illustrating selection of a grammar from the grammar store shown in Figure 1;

Figure 17 shows a flowchart for illustrating the performing of speech recognition by the speech server shown in Figure 1;

Figure 18 shows a flowchart for illustrating in greater detail the manner in which a user is advised if a machine cannot carry out the requested function;

Figures 19 to 24 show screens displayed to a user during operation of the system shown in Figure 1;

Figure 25 shows a block diagram similar to Figure 1 of a second embodiment of a system in accordance with the present invention;

Figure 26 shows a flowchart for illustrating another way of performing speech r cognition using the system shown in Figur 1 or Figure 25;

Figur 27 shows a block diagram of a modified personal computer suitable for use in the system shown in Figure 1 or Figure 25;

Figure 28 shows a flowchart for illustrating a modified way of performing speech recognition using the speech server of the system shown in Figures 1 or 25;

Figure 29 shows a block diagram illustrating another embodiment of a system in accordance with the present invention;

Figure 30 shows diagrammatically a user issuing instructions to a machine of the system shown in Figure 29;

Figure 31 shows a schematic block diagram of another embodiment of a system in accordance with the present invention;

Figure 32 shows a flowchart for illustratinginstallation of a machine onto the network of the system shown in Figure 31;

Figures 33 and 34 show different ways in which a user may be informed that the requested function cannot be carried out in the system shown in Figure 31;

Figures 35 and 36 are flowcharts for illustrating one mode of operation of the system shown in Figure 31.

[0014] Figure 1 shows by way of a block diagram a system 1 comprising a network N coupled to a number of different items of office equipment that will, in practice, be distributed throughout the building or buildings within which the network is installed. The network may be a local area network (LAN), wide area network (WAN), Intranet or the Internet. It should, however, be understood that, as used herein, the word network does not imply the use of any known or standard networking system or protocols and that the network may be any arrangement that enables communication between machines located in different parts of the same building or in different buildings. By way of example, Figure 1 shows a black and white photocopier 9, a colour photocopier 10, a facsimile (sometimes referred to as a "fax") machine 11, a digital camera 12, a personal computer 13, a multifunction machine 15 capable of copying, printing and facsimile functions and a printer 14 coupled to the network N. It will, of course, be appreciated that more than one of each of these different types of machines may be coupled to the network.

[0015] A speech server 2 is also coupled to the network N. The speech server 2 generally comprises a workstation or the like having a main processor unit 4 which, as known in the art, will include a CPU, RAM and ROM and a hard disc drive, an input device 21 such as, for example, a keyboard and a pointing device such as a mouse, a removable disc drive RDD 22 for receiving a removable storage medium RD such as, for example, a CDROM or floppy disc, and a display 25.

[0016] Program instructions for controlling operation of the CPU and data are supplied to the main processor unit 4 in at least one of two ways: 1) as a signal over the network N; and 2) carried by a removable data storage medium RD. Program instructions and data will be stored in the hard disc drive of the main processor unit 4 in known manner.

[0017] Figure 1 illustrates block schematically the main functional elements of the main processor unit 4 of the speech server 2 when programmed to operate in accordance with this embodiment of the present invention. Thus, the main processor unit 4 is programmed so as to provide an automatic speech recognition (ASR) engine 5 for recognising speech data input to the speech server 2 over the network N from any of the machines 9 to 13 and 15, a grammar store 8 storing grammars defining the rules that spoken commands must comply with and words that may be used in spoken commands, and an interpreter 7 for interpreting speech data recognised using the ASR engine 5 to provide instructions that can be interpreted by the machines 9 to 11, 14 and 15 to cause those machines to carry out the function required by the user with overall control of the speech server 2 being effected by a speech manager or processor 6. The speech server 2 also includes a machine identification (ID) store 3 storing data relating to each of the machines as will be described below.

[0018] Figure 2 shows block schematically the functional components of the copier 9 or 10. The copier 9, 10 comprises a main processor 20 coupled via an appropriate interface (not shown) to the machine functional circuitry which, in the case of the photocopier, consists essentially of optical drive circuitry 21, drum, exposure and development control circuitry 22 and paper transport circuitry 23. As is known in the art, the optical drive circuitry 21 controls illumination by the optical system of a document to be copied while the drum, exposure and development control circuitry 22 controls charging and exposure of the photosensitive drum and development of the resulting toner image. The paper transport circuitry 23 controls the transport of paper through the copier. The copier also includes a control panel 24 having manual controls for enabling a user to select the manner, type and number of copies to be produced. Thus, generally,the control panel will enable a user to select whether the copy is to be single or double-sided, collated or not collated etc. Also, in the case of a colour copier the control panel will generally enable the user to select whether a black and white or colour copy is required. The copier also has a display 25 for displaying messages and/or instructions to the user. Typically such a display is a liquid crystal display.

[0019] The above-described components shown in Figure 2 are present in known photocopiers. As shown in Figure 2, the photocopiers 9 and 10 shown in Figure 1 differ from known photocopiers in that they also include a microphone 26, a speech operation processor 27 and a network interface 28. The microphone 26 enables words spoken by a user to be converted into electronic speech data while the speech operation processor 27 is arranged to process the speech data for transmission over the network N via the network interface 28. The interface may be any known form of network interface, for example a TCP/IP socket where the network operates on the TCP/IP protocol.

[0020] Figure 3 shows a similar block schematic functional diagram of the facsimile machine 11. The facsimile machine 11 differs from the copier 9 or 10 in terms of its machine functional circuitry and these components of the facsimile machine ar known in the art. As shown block schematically in Figure 3, the machine function circuitry consists of optical driv circuitry 30 whichenables a document to be faxed to be scann d so as to convert the hard copy document into electronic form, transmit and receive circuitry 31 for transmitting and r ceiving facsimile messages, print head drive circuitry 32, a print carriage 33 and paper transport circuitry 34 for enabling a received facsimile m ssage to be printed out.

[0021] Figure 4 shows a similar block functional diagram of a digital camera. The digital camera differs from the photocopier 9, 10 in respect of its machine functional circuitry which consists essentially of exposure/zoom control circuitry 35, image capture and processing circuitry 36 which, for example, consists of an optical sensor and image compression circuitry, and an image or frame store 37 for electronically storing images captured by the image capture circuitry 36.

[0022] Figure 5 shows a block schematic functional diagram of the personal computer 13 shown in Figure 1. The personal computer 13 comprises a main processor unit 40 which, as known in the art, will include a CPU, RAM and ROM and a hard disc drive, an input device 41 such as, for example a keyboard and a pointing device such as a mouse, a removable disc drive RDD 42 for receiving a removable disc RD such as, for example, a CDROM or floppy disc and a display 45. Like the machines 9 to 12 described above, the personal computer also includes a network interface 28 for enabling connection to thenetwork N and a speech operation processor 27.

[0023] Program instructions and data may be supplied to the personal computer 13 connected to the network by supplying computer readable program instructions as a signal over the network N from, for example, another personal computer on the network or a remote device or by supplying computer readable instructions on a removable disc or storage medium RD.

[0024] Figure 6 shows a block schematic functional diagram of the printer 14. The printer 14 has a main processor unit 20 which controls operation of the machine functional circuitry which in this case consists of print head drive circuitry 51, print head carriage drive circuitry 52 and print transport drive circuitry 53. The circuitry 51, 52 and 53 enable the print head of the printer to print upon paper supplied to the printer in known manner. Where the printer is an ink jet printer, then the printer may also include recovery control circuitry 54 for, in known manner, causing the print head to be capped when not in use and for causing the print head to execute preliminary or idle discharge operations to clear any blockages of the ink jet nozzle.

[0025] The printer also includes a network interface 28 for enabling connection to the network N and a speech operation processor 55 arranged to receive instructions from the speech server 4 to cause the printer 14 to print in accordance with spoken instructions input, for example, to the personal computer 13 or the digital camera 12. Generally, the printer will not be provided with a microphone for speech command input because it is more convenient for the speech commands to be input from the machine from which the data to be printed is derived. Although not shown, the printer may also include a display.

[0026] In order to avoid accidental voice activation of any of the machines 9 to 12, they may be provided with a speech activation switch 29 shown in Figures 2 to 4 so that the machine is not responsive to speech input until the switch 29 is activated. A similar function may be provided in software for the personal computer 13 so that it is not possible for a user of the personal computer to send speech instructions to cause operation of a printer or fax machine until the user clicks on an appropriate icon displayed by the display of the personal computer.

[0027] The multifunction machine 15 will have all the functionality of the copier 9 or 10, facsimile machine 11 and printer 14.

[0028] As noted above, any known form of network N may be used. The speech operation processors 27, 55 and network interfaces 28 are provided as JAVA virtual machines using the JAVA programming language developed by Sun Microsystems Inc.

[0029] Each of the processors of the machines describedabove may be programmed by instructions stored in an associated storage medium and/or by signals supplied over the network N, for example.

[0030] The network also includes a look-up service 16 which contains a directory of all of the machines connected to the network together with their characteristics in a manner which enables a list to be extracted from the look-up service 16 of those machines which can perform a particular function. The JINI feature of the JAVA programming language is used so that the look-up service 16 is in the form of JINI look-up service and communication with the JAVA virtual machines forming the speech operations processors 27 is via a JINI service agent that requires only the interface to the machine to be standard so that the network protocols can be completely independent from the machines connected to the network.

[0031] Any known form of automatic speech recognition engine 5 may be used. For example, the automatic speech recognition engine supplied by IBM under the trade name "ViaVoice", the engine supplied by Dragon Systems, Inc. under the trade name "Dragon Naturally Speaking", the Microsoft speech recognition engine, the speech recognition engine produced by Nuance and that produced by Lernout and Hauspie. As will be understood by those skilled in the art, the speech manager 6 communicates with the automatic speech recognition engine 5 via a standard software interface known as "SAPI" (speech applications programmers interface) to ensure compatibility with the remainder of the system. In this case the Microsoft (Registered Trade Mark) SAPI is used.

[0032] The grammar store 8 stores rules defining the structure of spoken commands or instructions that can be input by a user and words that can be used in those commands. Although it would be possible to have a single grammar encompassing rules for commands for all of the types of machines coupled to the network N, in this embodiment separate grammars are provided for each different type of machine. Thus, the grammar store includes copy, fax and print grammar containing rules and words specific to copy, fax and print functions respectively. This has the advantage of limiting the choice of words available to the ASR engine 5 during speech recognition and so should reduce the possibility of misinterpretation of a word.

[0033] Different machines of the same type (for example different photocopiers) are of course capable of carrying out different functions. It is, however, desirable for machines of the same type to use the same grammar so that, for example, the photocopier and the colour photocopier 10 shown in Figure 1 are associated with the same copy grammar. One reason for this is that it enables a user to become rapidly familiar with the commands that can be used because they are the same for all machines of the same type (for example for all photocopiers). Another reason for this is that the user of a machine may not be fully familiar with the capabilities of that machine and

may ask the machine to carry out a function which is not available. If this happens and the command spoken by the user is not available within the associated grammar, then an error may occur because automatic speech recognition engines tend to return the closest possible match to a spoken word. Associating all machines of a given type with the same grammar should increase the possibility of recognition of commands input by a user and reduce the possibility of errors such as a user of the black and white photocopier 9 being presented with "collated" copies (when what he actually wanted were colour copies) because the grammar for the photocopier 9 did not recognise the word "colour". If, as will be described below, the black and white photocopier 9 and the colour photocopier 10 share a copy grammar, then a user requesting the black and white photocopier 9 to produce "four colour copies" can be advised that the black and white photocopier 9 is not capable of making colour copies.

[0034] As shown in Figure 7, the words included in the copy grammar include "copy", "copies" and synonyms such as "reproduce", "print" etc. The copy grammar also includes words relating to the manner in which the copying is to be carried out such as "single-sided", "double-sided", "reduced", "colour", "black and white", "collated", "A4", "A3", "stapled" etc.

[0035] The rules of the copy grammar will also look for a number or equivalent word which instructs the machine how many copies are required. These words may be stored in the copy grammar. However, it is more convenient for words shared by the copy, fax and print grammars ("shared words"), including numbers, to be incorporated in a separate, shared words grammar so that these shared words numbers can be accessed by any of the machines and it is not necessary to update the shared words grammar each time a new machine is connected to the network.

[0036] Figure 10 shows a list of words that may be included in a shared words grammar. These include introductory and closing phrases or words such as "please", "could you" etc. common to instructions for any of the machines shown in Figure 1. The shared words grammar will, of course, incorporate rules which determine whether these words are to be interpreted and, if so, how. Thus, for example, words such as "please", "could you" etc. will be ignored as not requiring interpretation. Although not shown, the grammar store may also include a shared grammar of natural numbers and a shared grammar of telephone numbers.

[0037] Figure 8 shows a list of words that may be included in the print grammar. These include words representing instructions such as "print", "hard copy ", "reproduce", words representing the manner in which printing should occur such as "black and white", "colour", "high resolution colour", "photoquality" etc. The print grammar will include pointers to the shared words grammar. The print grammar may also include a list of words that may be used to describe different types of printer such as "laser colour", "ink jet" etc.

[0038] Figure 9 shows a list of words and types of words that may be included in the fax grammar. These words include "fax", "send" and synonyms therefore plus words that may be used in connection with the desired destination of the facsimile message. For example, the fax grammar may include mnemonics, stored names or short codes for facsimile numbers, country and area facsimile codes etc. The fax grammar will also include pointers to the natural number and telephone number grammars where these exist.

[0039] The grammar store 8 will also include a JINI look-up service grammar that includes the words that may be used by a speaker to identify a printer desired to be used to print an image generated by a digital camera or personal computer connected to the network.

[0040] The above assumes that the user requires an immediate response from the machine. This need not, however, necessarily be the case and fax machines are available that enable a fax to be sent at a later timer date. Accordingly, the grammar store also includes a shared time/date grammar and the fax grammar will include rules as to how time/date inputs are to be interpreted. As shown in Figure 11, the time/date grammar includes words that may be used in connection with the time or date such as "am", "pm", "o'clock", "morning", "afternoon", "evening", "today", "tomorrow" and days of the week and months of the year. The time/date grammar will also include rules as to how input information is to be interpreted enabling times to be converted to a standard format such as a 24 hour clock. For example, the time/date grammar may include rules requiring the words "ten thirty" to be interpreted as "10:30" unless followed by the abbreviation "pm" in which case the words "ten thirty pm" will be interpreted as "22:30". The time/date grammar will also include rules defining which hours of the day constitute "morning", "afternoon", "evening" and for interpreting words such as "today", "tomorrow", "next Monday" as being specific days of the year defined in terms that can be interpreted by the internal clock and calendar of the speech server, network server and the machine itself.

[0041] Each grammar is a non-recursive context free grammar which determines the set of allowable spoken commands or utterances, together with attributes attached to each grammar rule which determine how the meaning is constructed. The attributes consist of fragments ofcomputer codes which are executed in dependence on the syntactic parse structure. In order to illustrate this, examples of rules that may be included in the time grammar are given below. In these examples, the computer programming language used is the Perl programming language.

```
Time =    {Time1:$time; Time2:$time; Time3:$time}⟨return ($time);⟩
Time1 =   {quarter⟨$m = 15;⟩; 1to29:m [minutes]} to 1to12:$h
          ⟨$hour=$h-1 % 12; $minute = 60-$m;
          $time = "$hour:$minute"; return($time);⟩
Time2 =   {half⟨$m = 30⟩; quarter⟨$m = 15⟩; 1to30:$m [minutes]} past 1 to 12:$h
          ⟨$hour = $h; $minute = $m;
          $time = "$hour:$minute";
          return($time);⟩
Time3 =   1to12:$h 1to59:$m
          ⟨$hour = $h;
          $minute = $m;
          $time = "$hour:$minute";
          return ($time);⟩
```

[0042] In these rules the curly brackets {} are used to group together alternative terms, square brackets [] denote optional terms and the angle brackets⟨⟩ contain Perl attributes. The Rule :$x construction is converted into Perl code which assigns the return value of "Rule" to the Perl variable $x. "1to12", "1to29", "1to30" and "1to59" are further rules (not shown) which accept natural numbers in the ranges 1 to 12, 1 to 29, 1 to 30 and 1 to 59, respectively, and return the corresponding value as a Perl variable. Together these rules accept time expressions

such as "quarter to ten", "fifteen minutes to ten", "half past ten", "quarter past ten", "fifteen minutes past ten" and "ten fifteen" and conv rt them all into a meaning of the form "hour:minute".

[0043] This grammar is converted into the SAPI recognition grammar used by the ASR engine 5 and a Perl file used by the interpreter 7. In this example the Microsoft (registered trade mark) SAPI is used. Upon compilation, parse tags are added to the SAPI recognition grammar. These parse tags label the individual items of each rule. In this example the PhraseParse method of the Microsoft grammar compiler (which is part of the Microsoft SAPI toolkit) is used to convert a recognised text string into a string of parse tags which indicate the syntactic parse structure of the recognised phrase. This string of parse tags is then passed to the interpreter 7 which executes the Perl code fragments in dependence on the parse structure.

[0044] Thus, the above grammar would be conv rted into the following SAPI 4.0 grammar format:

[⟨Time⟩]

[0045]

⟨Time⟩ = "(1" ⟨Time1⟩")"

⟨Time⟩ = "(2" ⟨Time2⟩")"

⟨Time⟩ = "(3" ⟨Time3⟩")"

⟨Time1⟩ = "(1" ⟨dummy1⟩")(2" to ")(3"⟨1to12⟩")"

⟨dummy1⟩ = "(1" quarter ")"

⟨dummy1⟩ = "(2" ⟨dummy2⟩ ")"

⟨dummy2⟩ = "(1" ⟨1to29⟩ ")"[opt]⟨dummy2b⟩

⟨dummy2b⟩ = "(2" minutes ")"

⟨Time2⟩ = "(1" ⟨dummy3⟩ ")(2" past ")(3"⟨1to12⟩ ")"

⟨dummy3⟩ = "(1" half ")"

⟨dummy3⟩ = "(2" quarter ")"

⟨dummy3⟩ = "(3" ⟨dummy4⟩ ")"

⟨dummy4⟩ = "(1" ⟨1to30⟩ ")" [opt]⟨dummy4b⟩

⟨dummy4b⟩ = "(2" minutes ")"

⟨Time3⟩ = "(1" ⟨1to12⟩ ")(2" ⟨1to59⟩ ")"

⟨1to12⟩ = ...

⟨1to29⟩ = ...

⟨1to30⟩ = ...

⟨1to59⟩ = ...

wherein the items enclosed in "" are the parse tags and the rules⟨dummy...⟩ are automatically generated by the pre-processor to make conversion to the SAPI 4.0 format possible.

[0046] In this example the phrase "ten minutes to two" would be converted into the following string of parse tags:

(1(1(2(1...)(2)))(2)(3...))

where the two sets of ellipses would be filled in by parse tags from the rules "1to29" and "1to12", respectively. Similarly "half past ten" would be converted to the following string of parse tags:

(2(1(1))(2)(3...))

with ellipses filled in from the rule "1to12".

[0047] The interpreter 7 then determines which bits of the Perl code to execute using the string of parse tags.

[0048] It will, of course, be appreciated that any known method may be used to determine the syntactic parse structure from a recognised phrase and that, moreover, the specific manner on which the rules are implemented and interpreted is not essential to the present invention. Any grammar structure that defines the set of allowable spoken commands or utterances and associates with each grammar rule attributes or data that determines how the meaning is constructed or extracted may be used.

[0049] The rules of the various grammars define the structures or formats of commands that are recognised by the grammars. Each grammar may include several different valid structures or formats. In this embodiment, the print grammar recognises the following two rules defining the structure or format of instructions:

Rule 1) [make] {a; and; 1-99} {copy; copies} [WhatPhrase] [CopyFeature];

Rule 2) {copy; duplicate} [WhatPhras] [HOW MANY] [CopyFeature].

[0050] In these rules square brackets are used to identify optional terms while curly brackets are used to group together alternative compulsory terms. Thus, in rule 1, the instruction format expected is an optional opening or introductory remark such as the word "make" followed by the word "a", "an" or a number between 1 and 99 representing the number of copi s followed by the word "copy" or "copi s" optionally follow d by a phrase that identifies the document to be copied ("WhatPhrase") optionally followed by a copy feature ("CopyFeature") identifying the manner in which the copying is to be carried out.

[0051] The individual components of this basic rule may be associated with sub-rules. For example, the "CopyFeature" is associated with the different ways of copying such as: "enlarged, reduced, collated, sort d, stapled, lighter, darker, colour, black and white etc." and each of these may in turn be associated with a sub-rule which defines the structure of a command relating to that particular copy feature.

[0052] As an example, if a person says "Make one copy of this document in black and white", the copy grammar will determine that the instruction complies with rule 1.

[0053] However, this copy grammar will not recognise the instruction "Please make a black and white copy" because the phrase "black and white" precedes the word "copy" and so the instruction does not comply with either rule 1 or rule 2. A modification of rule 1 may be stored in the copy grammar so as to enable a copy feature such as "colour" or "black and white" to precede the word "copy" or "copies".

[0054] In the case of copy features such as "enlarge" and "reduce", the copier needs further information to enable copying to be carried out, in particular the amount or degree of enlargement or reduction. Accordingly, the copy grammar will include appropriate sub-rules. Thus, for the copy feature "enlarge", then the sub-rule will define the valid structures and options for enlargement. For example, the sub-rule may expect the word "enlarge" to be followed by a paper size such as "A3" or "A4" or a word such as "maximum" or "max", a percentage or a multiplication. For example, the enlargement sub-rule may accept as a valid complete command "Enlarge this to A3" or "Enlarge this two times" but would require further instructions if the command spoken was simply "Enlarge this".

[0055] Copy rule 2 requires the instruction to start with the word "copy" or a synonym therefore such as "duplicate" optionally followed by a phrase representing what is to be copied itself optionally followed by a word or number identifying the number of copies and itself optionally followed by a word or phrase identifying a copy feature such as those set out above. For example, if a user simply says "Please copy this", then the copy grammar will identify this phrase as complying with copy rule 2 and will, as the default, determine that a single copy is required.

[0056] In a similar manner, the fax grammar includes abasic rule that expects a phrase of the type [please] {fax; send} [What Phrase] {mnemonic; shortcode; fax number} [time/date].

[0057] Thus, the fax grammar expects an instruction which may include an optional introductory word or remark such as the word "please" followed by a compulsory instruction in the form of the word "fax" (or a synonym such as the word "send") followed by an optional phrase identifying the document to be faxed such as the word "this" followed by a compulsory word, phrase or string of numbers identifying the destination of the fax followed optionally by words or numbers identifying the time and date of faxing. For example, this fax rule would accept as legitimate the phrase "Please fax this to Tom at 2pm tomorrow" if the word "Tom" is included in the mnemonics or names stored in the fax grammar. This fax grammar would not, however, accept a phrase of the type "Please fax tomorrow at 2pm to Tom" because the instruction order is wrong. However, the fax grammar may include an alternative rule for which this command would be valid.

[0058] The rules of the print grammar will follow the same basic principles as the copy grammar and may include the following basic rule: [please make] {a; an; 1 to 99;} {copy; hard copy; copies} [WhatPhrase] [CopyFeature] [PrinterIdentifier]. In this case the printer identifier will be a word identifying the printer such as "colour laser" ink jet etc. as shown in Figure 8 and the copyfeature may be for, example, colour, high resolution, photo-quality, reduced, enlarged etc. As in the case of the copy grammar there will be sub-rules for features such as enlarged, reduced etc.

[0059] The above printer rule is thus essentially similar to the copy grammar rule 1 but includes an additional optional component so that, for example, the instruction "Please print one copy black and white using the ink jet printer" is accepted as valid.

[0060] As explained above, each of the grammars may include a number of different valid formats or structures for entering commands. Although the definition of a number of valid commands, structures or formats means that not all commands input by a user will be recognised, the use of specified structures or rules for the input of commands assists in interpretation of the input instructions and reduces the possibility of ambiguity so that, for example, differentiation can be made between the numeral "2" and the word "two" on the basis of location within the input command. This should also facilitate identification and separation of numbers such as telephone numbers from numbers identifying times or dates or numbers of copies, for example.

[0061] The manner in which a piece of office equipment such as one of the machines 9 to 15 shown in Figure 1 is coupled to the network N will now be described with reference to the flowchart shown in Figure 12.

[0062] Because the JAVA/JINI platform is being used, a new machine is automatically located by the speech server on installation at step S1 and the machine ID and data identifying the associated grammars is stored in the machine ID store 3 at step S2. The new JINI compliant machine is also automatically registered with the JINI look-up service 16 so that the details of the new machine and the functions it is capable of performing are stored in the look-up service at step S2a. Figure 13 shows an example of the type of data that may be held in the look-up service 16 for the copier 9.

[0063] The speech server 2 then determines whether all of the associated grammars identified in the machine identification data are already available in the grammar store 8 at step S3. If the answer is no, then the speech server 2 communicates with the newly installed machine requesting the newly installed machine either to supply a copy of the associated grammar(s) stored in the memory of its speech operational processor 27 or to provide the

speech server 2 with a network or an Internet or worldwide web address from which the speech server 2 can download the grammar(s) (step S4 in Figure 7). The speech server 2 will also check whether the newly installed machine has information regarding updates for existing grammars and, if so, will download these to replace the older existing versions.

[0064] The procedure described above with reference toFigure 12 will be the same regardless of the type of JINI compliant machine being installed on the network although, of course, the information in the grammars will vary dependent on the type of machine.

[0065] Operation of the system shown in Figure 1 will now be described with ref rence to Figures 14 to 24.

[0066] Figures 14 and 15 are top level flowcharts illustrating, respectively, the functions carried out by a machine having a copying or facsimile transmission function (machines 9 to 11 and 15 in Figure 1) at which a user is inputting instructions (the "originating machine") and the corresponding functions carried out by the speech server 2.

[0067] At step S6 in Figure 14, the main processor of the originating machine checks to see if spoken operation of the machine has been instructed. If the answer is no, the machine responds to instructions input manually using the control panel 24 (see for example Figures 2 to 4) at step S7 and, if the instruction is complete at step S16, proceeds to cause the machine to carry out the required function at step S18.

[0068] If, however, the main processor of the originating machine determines at step S6 in Figure 14 that its speech operation activation switch 29 has been activated, then the speech operation processor 27 will rec ive, process and store speech data input via the microphone 22 ready for transmission to the speech server 2 via the network N in accordance with the network protocol (step S8).

[0069] Before sending the received speech data, the speech operation processor 27 needs to establish communication with the speech server 2. This is illustrated by steps S9 and S19 in Figures 14 and 15, respectively. Generally, this communication involves the originating machine sending over the network a message to the speech server 2 identifying itself and requesting permission to send speech data to the speech server 2 (step S9 in Figure 14). In response to receipt of such a request, the speech manager 6 of the speech server 2 identifies the machine making the request and in response sends a message either granting or denying the request (step S19 in Figure 15).

[0070] Once the originating machine receives from the speech server 2 a message granting its request to send speech data, then the speech operation processor 27 of the originating machine sends the speech data (step S10 in Figure 14) which is received by the speech server 2 at step S20 in Figure 15. Initially the speech manager 6 will use a general start-up grammar for speech recognition. This start-up grammar may consist of all available grammars or just a set of rules and words that may be used at the start of a command. Having identified the originating machine, the speech manager 6 selects, using the data stored in the machine ID store 3, thegrammars associated with the originating machine from the grammar store 8 at step S21 and then controls the ASR engine 5 to perform speech recognition on the received speech data in accordance with the selected grammars at step S22 in Figure 15. Once the ASR engine 5 has completed the speech recognition process, the speech manager 6 causes, at step S23 in Figure 15, the interpreter 7 to interpret the recognised speech to produce language independent device control commands that can be used by the speech operation processor 27 of the originating machine to supply to the main processor 20 of that machine control commands emulating those that would have been supplied if the command had been entered manually using the control panel 24.

[0071] The speech manager 6 then communicates with the originating machine, sends the device control commands produced by the interpreter 7 to the originating machine at step S24 and then returns to step S19 awaiting further communication from machines on the network N.

[0072] If no device control commands or instructions are received by the originating machine at step 11, the machine waits for a predetermined period of time at step S12 in Figure 14 and, if no response is received from the speech server 2 after that predetermined period of time, displays to the user at step S13 in Figure 14 an error message. The speech operation processor 27 then returns to point A in Figure 14 and awaits further manual or spoken input by the user.

[0073] Once device control commands have been received at step S11 in Figure 14, the speech operation processor 27 processes these device control commands and provides to the main processor 20 control commands emulating those that would have been provided if the instructions had been input via the control panel. The main processor 20 then compares the received control commands with the control commands associated with that machine at step S14. If the received control commands are not recognised, the main processor 20 advises the user accordingly at step S15 in Figure 14 and the speech operation processor 27 and main processor 20 return to point A in Figure 14 awaiting further input from the user.

[0074] Because the system 1 shown in Figure 1 includes the look-up service 16, if the received control commands are not recognised, the machine can request (step S151 in Figure 18) the look-up service 16 to search its database for a machine that can perform the requested function and, when the look-up service 16 returns the requested information, display a message to the user on the display 25 saying "This machine cannot do .... However machine .... can." (step S152 in Figure 18). For example the message "This machine cannot do colour copying. However machine No. 10 can." may be displayed if the user has requested colour copies from copier 9. This allows the user the options of moving to the machine that can carry out the required function or of modifying his original request so that the current machine can carry it out.

[0075] If the main processor 20 identifies the control commands at step S14, then the main processor 20 will proceed as if the control commands had been input using the control panel 24. Thus, at step S16 in Figure 14, the main processor 20 will determine whether the instruction input by the user is complete. For example, where the originating machine is the photocopier 9 and the user has requested an enlarged copy but has not specified the degree of enlargement, then the main processor 20 will determine at step S16 that the instruction is incomplete and will display to the user at step S17 a message requesting further instructions. In this example, the message will request the user to input the degree of enlargement. The main processor 20 and speech operation processor 27 will then return to point A in Figure 14 awaiting further manual or oral input from the user.

[0076] It will be appreciated from the above that instructions can be input by a user to the machines 9, 10 or 11 using a combination of manual and orally entered commands so that, for example, when further information such as the degree of enlargement is requested that can be input orally or manually using the control panel 24. The reason for this is that, as shown in Figure 14, whenthe speech operation processor 27 and main processor 20 of the

machine 9, 10 or 11 has either carried out the requested function (step S18) or displayed a message to the user (step S13, S15 or S17), the machine returns to point A in Figure 14 awaiting further manual or oral instructions.

[0077] Although the speech server 2 may access the entirety of the grammars associated with the originating machine 9, 10 or 11 each time the speech server 2 receives speech data from that machine, it is more efficient and accurate for the speech operation processor 27 and speech server 2 to operate in different dialogue states dependent upon the nature of the instruction input by the user so that, for example, each time the originating machine prompts the user to input further instructions at step S17 in Figure 14, the speech operation processor 27 enters a different dialogue state and, when further instructions are input by the user, transmits this different dialogue state to the speech server 2 to enable the speech manager 6 to access only the grammars or grammar portions relevant for that dialogue state. This will now be described in detail taking as a first example the case where a user wishes to use the multifunction machine 15 to produce an enlarged copy.

[0078] As described above with reference to Figure 14, the machine 15 will normally be in an initial state awaiting input of instructions from a user. In this state, the main processor 20 of the multifunction machine 15 will cause its display 25 to display to the user a message requesting the user to select one of the available functions. Figure 19 shows a typical message on the screen 25A of the display 25. As shown, the display may list the available functions.

[0079] In this example, the user wishes to use the copying function of the machine 15 and so, after operating the speech activation switch 29, inputs via the microphone 29 the command "copy" or a phrase incorporating that word such as "I wish to copy".

[0080] The speech operation processor 27 then processes and stores the entered speech data at step S8 in Figure 14 ready for transmission on the network and communication between the machine 15 and the network server 2 in preparation for sending of the speech data to the speech server is carried out at steps S9 and S19 in Figures 14 and 15 respectively.

[0081] Once the machine 15 has successively received from the speech server 2 permission to send the speech data, the speech operation processor 27 of the machine 15 sends the speech data to the speech manager over the network and in accordance with the network protocol together with information identifying the dialogue state of the machine. Initially, the machine will be in a start-up dialogue state (step S121 in Figure 16) and accordingly the speech manager 6 will select at step S121a in Figure16 the general start-up grammar that, as noted above, may consist of all the grammars stored in the grammar store 8.

[0082] The speech manager 6 then causes the ASR engine 5 to perform speech recognition on the received speech data using the selected grammar to extract the meaning from the received speech data (step S22 in Figure 15 and step S221 in Figure 17) and then checks whether the results of the speech recognition include unidentified words or phonemes at step S222. If the answer is yes at step S222, then the speech manager 6 determines that the input words were incomplete and sends to the originating machine a message that the words were not recognised together with instructions to cause the originating machines to display a message to the user indicating that the instructions were not recognised (step S223 in Figure 16). The speech manager 6 then returns to point B in Figure 15 awaiting further communication requests from machines coupled to the network.

[0083] Assuming that the answer at step S222 in Figure 17 is no, then the speech manager controls the interpreter 7 at step S23 in Figure 15 to interpret the results of the speech recognition process to produce language independent device control commands which are then sent to the originating machine, in this case the multi-function machine 15.

[0084] In this example, the user has input to the multi-functionmachine 15 a command indicating that the copy function is required and accordingly the ASR engine 5 will have recognised the word "copy" and the interpreter 7 will interpret this as a language independent command for the machine 15 to enter the "copy" dialogue state. When this command is received by the multi-function machine 15, the speech operation processor 27 and main processor 20 will cause the machine 15 to enter the copy dialogue state and to display on the screen 25 a message requesting further instructions from the user. Figure 20 shows the screen 25A of the display illustrating a typical message: "what type of copy do you want?".

[0085] The multi-functional machine 15 is now in a dialogue state in which it expects the user to input a command that complies with the grammar rules of the copy grammar and the shared words and time/date grammar. When such a command is received from the originating machine (step S20 in Figure 15), the speech manager 6 determines from the information accompanying the speech data that the originating machine is in a copy dialogue state (step S122 in Figure 16) and accordingly selects the copy, shared words and time/date grammars at step S123 in Figure 16. The speech manager 6 then causes the ASR engine 5 to perform speech recognition (step S22 in Figure 15) using the selected copy, shared words and time/date grammars and, assuming the answer at step S222 in Figure 17 is no and there are no unidentified wordsor phonemes, then the speech manager 6 controls the interpreter 7 at step S20 in Figure 15 to interpret the results of the speech recognition process to produce language independent device control commands which are then sent to the originating machine. As will be explained below, these device control commands may enable the originating machine to carry out the required function or may cause the originating machine to enter a dialogue state subsidiary to the copy dialogue state in which it is expecting further instructions in relation to the copy command input by the user.

[0086] It will, of course, be appreciated that if the user had input to the multi-function 15 a command indicating that the fax or print function is required so that the answer at step S124 or S126 in Figure 16 was yes, then the fax, shared word and time/date grammars or the printer, shared words and time/date grammar would have been selected at step S125 or S127 instead of the copy, shared words and time/date grammar. If at step S128 in Figure 16, the speech manager 6 determines from the information accompanying the speech data that the originating machine is in a dialogue state other than the top level or basic copy, fax or print dialogue state, then the speech manager 6 will select at step S129 the grammar or grammar portions relevant to that dialogue state. Such a dialogue state would arise where, for example, the originating machine requires additionalnstructions to complete a copy function as will be described in greater detail below.

[0087] Two different examples of copy commands will be described below.

[0088] In a first example, the user inputs the command "Please make a black and white copy" in response to the screen 25A shown in Figure 20. When this command is received by the speech operation processor 27, then steps S8 to S10 of Figure 14 are carried out as described above and the speech operation processor waits at step S11 in

Figure 14 for device control commands from the speech server 2. The speech server 2 processes the received speech data as described above with reference to Figure 15 using the copy, shared words and time/date grammars and returns to the machine device control commands to instruct the machine to make one black and white copy of the document placed on the copy glass of the machine 15 by the user. Upon receiving these device control commands, the speech operation processor 27 determines that the machine 15 is, in this example, capable of carrying out the requested function and accordingly supplies to the main processor 20 instructions emulating those that would have been supplied if the user had used the control panel 24 to input the instructions.

[0089] In the above first example, the instructions supplied in response to the prompt shown in Figure 20 were complete. The second example to be described represents a case where the instruction supplied by the user is incomplete. In this example, in response to the prompt shown in Figure 20, the user inputs the command "Please make a black and white copy enlarged". This input speech data is communicated to the speech server 2 and processed as described above with reference to Figures 14 and 15 so that the speech server 2 sends back to the originating machine 15 language independent device control commands that the speech operation processor 27 processes to supply to the main processor 20 instructions that emulate the instructions that would be received by the main processor 20 if the user had manually requested a black and white enlarged copy. The main processor 20 determines that the user has not specified the degree of enlargement required and accordingly displays on the screen 25A of display 25 a message requesting the user to enter the degree or amount of enlargement required. A typical message is shown in Figure 21. The speech operation processor 27 is now in an auxiliary dialogue state in which it is expecting input by the user of spoken or manual commands in response to the prompt shown in Figure 21.

[0090] In this example, it is assumed that the user inputs the further commands using the microphone 22. These auxiliary instructions are then communicated to the speech server 2 as speech data accompanied by information indicating the auxiliary dialogue state of the machine15. The speech data is processed as described above with reference to Figures 14 to 18. However, in this case, the speech manager 6 determines at step S128 in Figure 16 that the originating machine is in an auxiliary dialogue state, identifies this dialogue state from the information supplied by the machine 15 and selects the appropriate portions of the copy grammar at step S129 in Figure 16, in this example the copy feature dialogue state "enlarge". The speech manager 6 thus selects the portion (sub-grammar) of the copy grammar which relates to enlargement and includes rules and words determining the manner in which enlargement may be specified by the user. For example, the enlargement section of the copy grammar may enable input of commands such as "to A3", "as big as possible", "x times" where x is an integer, or "y%" where y is a number greater than 100. The speech manager 6 then controls the ASR engine 5 to interpret the speech data received from the machine 15 using this "enlargement" sub-grammar.

[0091] It will thus be seen that the speech manager 6 selects for the speech recognition processing only the grammars or portions of the grammars that are related to the dialogue state of the machine as identified by the data accompanying the speech data to be processed. This restricts the selection of words and phrases available to the ASR engine 5 which should increase the efficiency of the speech recognition process and also should reduce the likelihood of ambiguous or incorrect results because words that are not relevant to the current dialogue state of the originating machine will not be contained in the grammar or sub-grammar accessible to the ASR engine 5.

[0092] It will be appreciated that where the originating machine is the copier 9 or 10 rather than the multifunction machine 15, that the first screen shown to the user will be a screen corresponding to the screen 25A shown in Figure 20 (not Figure 19) and that the speech manager 6 will generally determine that the dialogue state is the copy dialogue state (step S122 in Figure 16) from the machine identification.

[0093] Where the originating machine is the fax machine 11, then the screen 25A of the display will initially display to the user a message along the lines shown in Figure 22 requesting the user to identify the destination to which the fax is to be sent. In this case, when the user responds by operating the speech activation switch 29 and speaking into the microphone "Please fax to Tom", then the speech manager 6 will determine at step S124 in Figure 16 that the originating machine is a facsimile machine and will select the fax, shared words and time/date grammars, process the received speech data using the ASR engine 5 with these grammars and then interpret the results using the interpreter 7 as described above with reference to Figures 14 and 15 so as to return to the fax machine 11 device control commands instructing the fax machine to fax the message to the facsimile number associated with the mnemonic "Tom". In this example, the fax machine is capable of delayed transmission of facsimile messages and accordingly the main processor 20 will determine that the instruction input via the user is incomplete and will display on the screen 25A a message to the user requesting them to specify when they want to send the fax message (see Figure 23). The speech operation processor 27 then enters an auxiliary dialogue state in which it is expecting a response identifying a time/date and any further speech input by the user is communicated to the network along with data identifying the speech data as being related to a "time/date dialogue state". Upon receipt of this speech data, the speech manager 6 will determine that the speech data should be interpreted using the time/date grammar and so will select only this grammar for use by the speech recognition engine 5. Once the meaning has been extracted from the speech data, the speech manager 6 controls the interpreter 7 to generate language independent device control commands corresponding to the time/date speech data and supplies these to the originating machine in the manner described above with reference to Figures 14 to 18.

[0094] The above description assumes that the machine originating the instructions is also the machine that will carry out the instructed function (that is one ofmachines 9 to 11 or 15 in Figure 1). This will not be the case where the user is issuing instructions via the microphone of the digital camera or personal computer 12 or 13 to cause printing via the printer 14 (or the printing function of the multifunction machine 15).

[0095] As an example, assume that after activating the speech activation switch 29, a user of the digital camera 12 (or personal computer 13) says "Please print this in photoquality".

[0096] Upon receipt of the speech data representing this message (step S20 in Figure 15), the originating machine will be identified using the start up grammar as the digital camera 12 and the speech manager 6 will select the print, shared names, time/date grammars and JINI look up service (steps S126 and S127 shown in Figure 16). The ASR engine 5 will then, under the control of the speech manager 6, perform speech recognition in accordance with the selected grammars. If the user does not specify the type of print then the prompt shown in Figure 24 may be displayed and further data awaited.

[0097] As noted above, the printer grammar optionally expects the information in the speech data identifying the type of printer. Where the network incorporates only one printer 14 as shown in Figure 1, then the speech

manager 6 will cause the JINI service object associated with that print r to be downloaded to the camera (or personal computer) as the default. However, where then twork includ s a number of different printers and the received speech data does not identify the printer then th speech operations processor 27 of the printer will cause the main processor 20 to display on the display 25 of the digital camera 12 (or the display of the personal computer 13) a message saying "Please select printer" (step S17 in Figure 14). The speech operations processor 27 will then enter a print r select dialogue state and this dialogue state will b transmitted to the speech server together with any further spoken instructions input by th user so that when these further spok n instructions are received by the speech server, the speech manager will select the JINI look-up servic grammar at step S129 in Figure 16 and will cause an instruction to be sent to the JINI look-up servic 16 to cause the service object for the printer specified in the further spoken instructions to be downloaded to th camera 12.

[0098] If the downloaded JINI service object determines that the information required to carry out printing is complete it sends the printer control commands together with the data to be printed to the required printer. If however the JINI service object determines at step S14 that the associated printer cannot produce the type of print required by the user, then the JINI service object will interrogate the JINI look-up service 16 to determine whether there is a printer available on the network thatcan print in accordance with the user's instructions and will advise the user accordingly (see steps S151 and S152 in Figure 18).

[0099] Thus, in the present embodiment, because the speech operation processors 27 constitute JAVA virtual machines using the JINI facility, the fact that the machine at which the instructions are originated is not the machine at which the instruction will be carried out makes very little difference to the operation of the system. This is because the JINI service object for the or the requested printer will be downloaded to the JAVA virtual machine of the digital camera 12 or personal computer 13 so that the digital camera 12 or personal computer 13 can carry out steps S14 and S16 in Figure 14 without the device control commands received from the speech server 2 having to be sent to the printer. Once the digital camera 12 or personal computer 13 determines using the downloaded JINI service object that the print instruction is complete at step S16 in Figure 14, then the JAVA virtual machine constituting the speech operation processor 27 of digital camera 12 or personal computer 13 will send the print control commands together with the data to be printed to the required printer 14 at step S18 in Figure 14.

[0100] Each of the above-described examples assumes that the user requires the requested function to be carried out immediately. This need not, however, necessarily bethe case and the user may require, for exampl , that transmission of a facsimile message or printout of a document be deferred until a later time or date. For this reason, the fax and printer grammars both include optional time/data components determined in accordance with the rules stored in the common time/date grammar.

[0101] The time/date grammar is arranged to cooperate with the system clock and calendar of the network s rver and the speech server and, as mentioned above, uses built-in rules to interpret spoken words in accordance with the system clock and calendar.

[0102] Accordingly, if the user instructs the facsimile machine 11 to "Fax this to Tom at 10am tomorrow", then this instruction will be interpreted in the same manner as the instruction "Fax this to Tom". However, in addition, the words "at 10am tomorrow" will be interpreted by use of the fax and time/date grammars by the speech interpreter 7 to produce device control commands that will cause the fax machine 11 to defer transmission of the fax until the fax clock time that corresponds to "10am tomorrow", provided that the machine ID store indicates that the facsimile machine 11 can perform delayed facsimile transmission.

[0103] If the speech manager 6 determines from the machine ID store 3 that the facsimile machine 11 does not have the facility for delayed or deferred transmission of facsimile messages, then a message to this effect may be displayed to the user at step S101 in Figure 16. Alternatively, the speech server 2 may provide a data store where the facsimile message to be transmitted can be stored until the required time and then returned to the facsimile machine 11. It will, of course, be appreciated that a similar approach could be used to defer printing of documents until a specified time so that, for example, a user at a personal computer remote from the required printer can defer printing of the document until he can be present at the desired printer. Where this feature is available, then printing may be deferred until the user inputs a further command along the lines of "Please print job number .....". If the printer itself is provided with a microphone and the same speech operation processor software as the copying and facsimile machines 9 to 11, then this deferred print command could be input by the user at the printer so that the user can ensure that the document is not printed until he is actually present at the desired printer and has instructed its printing. This may be of particular advantage where the printer is located at some distance from the originating machine (digital camera or personal computer) and, for example, the user does not wish other people to view the document being printed.

[0104] Figure 25 illustrates another embodiment of a system 1a in accordance with the present invention. The system 1a differs from the system 1 shown in Figure 1 in thathe speech server 2 also includes in the grammar store 8 a voice macro store 8a which includes voice commands defining specific instructions or instruction sequences. If these voice macros are specific to a type of machine, for example a copier, then they may be included in the copy grammar. However macros common to two or more types of machines will be stored in a shared voice macro store.

[0105] Where the machine is a copier, then voice macros such as "monthly report format", "house style" etc. may be defined at the speech server 2 so that every time a user requires a document to be in monthly report format or house style this can be achieved simply by the user saying "monthly report format" and the speech manager 6 will determine (when these words have been identified by the ASR engine 5) from the voice macro grammar what copier functions correspond to the voice macro "monthly report format", for example the voice macro "monthly report format" may correspond to the instruction 20 black and white copies, double-sided, collated and stapled. The speech interpreter 7 will then produce the device control commands required for the speech operation processor 27 of the originating copier 9 or 10 to cause the copier to produce copies in accordance with the required monthly report format. This central or shared storage of voice macros means that modifications or updates of standard copying styles can be effected and it is not necessary for each individual user to know thedetails of the current format but simply that the format exists. Voice macros may similarly be stored for different printing styles for different documents defining, for example, the print quality and number of copies and voice macros may also be provided for the facsimile machine so that where documents are frequently faxed to a group of people at different facsimile numbers, the voice macro to "fax to set 1" will be interpreted as requiring the same document to be faxed to each of

the facsimile numbers in set 1.

[0106] The functioning of the apparatus shown in Figure 25 differs from that described with reference to Figure 1 in that the speech manager 6 will select the voice macro grammar as well as the basic (copy, fax, print) grammar. Of course, where the voice macro grammar is embedded in the corresponding one of the other grammars then the voice macro grammar will automatically be selected. Otherwise the operation of the syst m is the same.

[0107] In the above-described embodiments, a single ASR engine 5 which is not trained to any particular person's voice or speech is used. The speech server 2 may, however, store a number of different ASR engines each trained to a particular person's voice or speech. In this case, the grammar store 8 would also include a voice grammar associating each ASR engine with a particular user and including words and grammar rules that may be used to identify that user. In such a case, where, foexample, the user wishes to produce copies at the photocopier 9, then the user may say "This is John. Please make two copies double-sided". In this case once the speech data is received by the speech manager 6, the speech manager 6 will select the grammar for the originating machine as described above (step S21 in Figure 15) but will also select the voice grammar. The speech manager 6 then causes, at step S900 in Figure 26, a default ASR engine to perform preliminary speech recognition with the help of the voice grammar so as to determine whether any of the words spoken by the speaker (the word "John" in the above example) identify that speaker as a person for whom the speech server 4 has access to a trained ASR engine.

[0108] If at step S901 in Figure 26 the speech manager 4 determines that the speaker has been recognised, then the speech manager selects at step S903 the ASR engine associated with that particular speaker and then performs speech recognition using the selected ASR engine at step S904. If the answer at step S901 is no and th speaker is not recognised, then a default ASR engine is selected at step S902 and then speech recognition is performed using the default ASR engine at step S905. Once either speech recognition engine has completed its task, the speech manager 6 determines whether there are any unidentified words or phonemes at step S222 and, if so, sends the required message at step S223 and then returnsto point B in Figure 15. Otherwise the speech manager 6 proceeds to step S23 in Figure 15.

[0109] The trained ASR engines 5 may be stored at the speech server 4 or may be located on individual personal computers connected to the network. Figure 27 shows a block schematic diagram of a modified personal computer 13a which also includes software defining a trained ASR engine 5a. Where the trained ASR engines 5a are located on personal computers 13a, then, of course, the voice grammar will include the necessary information for the speech manager 6 and interpreter 7 to download or access the required ASR engine over the network N.

[0110] Where the network is such that individual personal computers, copiers, fax machines and digital cameras are each primarily used by a particular different individual, then this information may be included in the machine ID store 3 so that the speech manager 6 can identify the most likely user of the machine from the identity of th machine being used. In such circumstances, of course, step S900 in Figure 26 will be omitted and step S901 of recognising the speaker will be carried out on the basis of the identified machine and its association with a particular individual rather than on the words spoken by the individual user.

[0111] Even where the individual user of a machine cannot be identified, using a number of different ASR engines and comparing the results of the speech recognitionperformed by those different ASR engines may increase the reliability of the speech recognition. As described above, these different ASR engines may be located at the speech server or distributed between different personal computers connected to the network. Figure 28 shows a flowchart for illustrating the process of performing speech recognition where a number of different ASR engines may be available.

[0112] In this case, once the speech manager 6 has received the speech data and has selected the appropriate grammars (for example the copy and shared words grammars for a photocopier), the speech manager 6 checks at step S906 in Figure 28 whether there are ASR engines available on personal computers connected to the network. If the answer at step S906 is no, then the speech manager 6 selects the speech server ASR engine 5 as the default at step S907.

[0113] If the answer at step S906 is yes, then the speech manager 6 selects all currently idle ASR engines connected to the network at step S908 and uses, at step S909, each of these ASR engines to perform speech recognition on the received speech data. The speech manager then compares, at step S910, the results of the speech recognition carried out by each of the selected ASR engines and selects, at step S911, the most likely result. This may be achieved by, for example, using a voting scheme so that the most commonly occurring wordor words in the speech recognition results are determined to be the most likely spoken words or a confidence scheme based on the confidence scores provided as part of the ASR engines may be used. Steps S222 and S223 then proceed as described above with reference to Figure 17.

[0114] Each of the embodiments described above requires the machine to which spoken instructions are to be input to be provided with its own microphone. Figure 29 illustrates another embodiment of a system 1b in accordance with the present invention wherein the speech server 2 is coupled to the internal exchange 17 of a digital enhanced cordless telecommunications (DECT) internal telephone network.

[0115] In this embodiment, a user may input instructions to one of the machines 9, 10, 11, 12 or 13 using his DECT mobile telephone. Figure 30 illustrates schematically a user U instructing the copier 9 to produce copies using a DECT mobile telephone T.

[0116] Each mobile telephone on the DECT exchange will normally be used only by a specific individual. That individual may, however, be located adjacent to any one of the machines coupled to the network when the speech server 2 receives instructions via the DECT exchange 17. It is therefore necessary for the speech server 2 to be able to identify the location of the DECT mobile telephone T (and thus the user) so that the speech server 2 can determine which of the machines 9 to 13 or15 is receiving instructions from the user. It may be possible to determine the location of the mobile telephone from the communication between the mobile telephone and the DECT exchange 17. However, in this embodiment, each of the machines coupled to the network is given an identification. Thus, as shown in Figure 30, the copier 9 carries a label identifying it as "copier 9".

[0117] In this embodiment, the grammar store 8 also includes a DECT grammar 8b which includes rules specifying the structure of phrases that the user can input using his DECT telephone to identify the machine at which he is located together with the words that may be used to input those instructions. For example, the DECT grammar may allow phrases such as "I am at copier number 9" or "This is copier number 9".

[0118] Operation of the system 1b shown in Figure 29 differs somewhat from that shown in Figure 1. Thus, in

this embodiment, the user issuing instructions via his DECT telephone first of all has to identify the machine at which he is located. Accordingly, the initial dialogue state for instructions issued using the DECT telephone will be a "DECT dialogue state" (that is one of the other dialogue states at step S128 in Figure 16) and accordingly initially the speech manager 6 will select the DECT grammar 8b so that the ASR engine 5 can interpret the words spoken by the user in accordance withthe DECT grammar 8b so as to enable the speech manager 6 to identify the originating machine. Once the originating machine has been identified, then the speech server 2 will send control commands to the machine at which the DECT telephone is located in the manner described above with reference to Figures 19 to 24 so that the display 25 of the originating machine displays to the user a message prompting the user to input further information. Where, as shown in Figure 30, the user is determined to be located at the copier 9, then the screen shown in Figure 20 will be displayed to the user. The user may then input further instructions using the DECT telephone or the control panel of the machine and processing of those instructions will proceed as described above with reference to Figures 14 and 15.

[0119] Although providing prompts and requests for further instructions to the user via the display is straightforward, it may be more convenient for the user to receive prompts or further instructions via his DECT telephone. Accordingly in a modified form of the system 1b the speech manager 6 is associated, as shown in Figure 29, with a speech synthesis processor 60. In this modified embodiment, when the speech manager 6 determines that a prompt needs to be sent to the user, instead of sending instructions over the network N to the speech operation processor of the originating machine to cause the associated main processor to display the appropriate message on its display 25, the speech manager 6 sends the text of the prompt to the speech synthesis processor 60 which generates the equivalent spoken form of the text and then transmits this to the user via the DECT exchange 17. The use of the DECT mobile telephone system to send prompts and requests for further instructions to the user has the advantage that it is not necessary to use a display present on the originating machine and, for example, the system may be used where the originating machine has no or a very small visual display, as may be the case for the digital camera 12.

[0120] Operation of the system 1b where prompts and requests for further instructions are sent to the user via the DECT telephone system is similar to that described above except that messages such as those shown in Figures 19 to 24 will be given to the user orally over the telephone system rather than visually via the display of the originating machine.

[0121] The system shown in Figure 29 may be adapted for use with a conventional fixed line internal telephone exchange and that, if each of the machines on the network is located adjacent to a specific fixed location telephone, then it will be possible to identify the location of a user automatically so that it is not necessary for the user to identify the machine from which he is issuing instructions.

[0122] Although the system 1b shown in Figure 29 enablesthe need for microphones and speech processing software at the machines 9 to 13 and 15 to be avoided, it is still necessary for these machines to incorporate a modified form of the speech operation processor 27 to enable device control commands to be received from the speech server 7 and processed to provide the main processor 20 with commands that emulate the commands that would have been provided if the manual control panel 24 had been used. The need for the speech operation processors 27 could be removed entirely if the interpreter 7 was programmed so as to supply to each of the machines 9 to 12 and 15 device control commands that directly emulate the device control commands that would have been supplied to the machine via the control panel. This would mean that the only modification required of these machines to enable speech controlled operation would be to modify the main processor and its operating software so as to enable it to take control commands from the network as well as the control panel.

[0123] The system 1b shown in Figure 29 enables oral prompts and requests for further information to the user. The system 1 shown in Figure 1 and the system 1a shown in Figure 25 may also be modified to enable oral prompts to the user by providing each of the machines with a loudspeaker and providing speech synthesis software either in the speech operations processor 27 of the individual machines or at the speech server 2 byincorporating in the speech server 2 a shared speech synthesis processor 60 as shown in Figure 29.

[0124] The system 1b has particular advantages where ASR engines trained to the voice of particular users are available to the speech server 2. Thus, where the DECT telephone system is used to input instructions, then it will not generally be necessary to perform the preliminary speech recognition step S900 shown in Figure 26 because the speech manager 6 will be able to identify the speaker from the identity of the DECT mobile telephone because, as mentioned above, each DECT mobile telephone will be associated with a particular user.

[0125] In each of the embodiments described above, the speech operations processor 27 forms a JAVA virtual machine and uses the JINI facility of the JAVA platform. It is, however, not necessary to use the JAVA platform and other platforms that provide similar functionality may be used.

[0126] The present invention may also be implemented using operating software and platforms that do not provide the functionality of the JAVA and JINI systems described above. In such cases, it should still be possible to use the look-up service 16 described above but entry of information into this look-up service may need to be carried out by, for example, a manager of the network.

[0127] As described above, the JINI service objects enable the digital camera 12 or personal computer 13 to determine whether the printer 14 is capable of acting in accordance with the device control commands received back from the speech server.

[0128] Figure 31 illustrates a modified system 1c which does not have the functionality of the JINI service object. This system also does not include the look-up service 16. Otherwise, this embodiment is similar to that shown in Figure 1.

[0129] Figure 32 illustrates the installation of a new machine onto the network N of the system 1c. In this case, installation of a new machine at step S1a is not automatic and it is necessary for the network manager to install the new machine onto the network in conventional manner and then to ensure at step S2 that the machine is provided with or supplies to the speech server 2 a machine identification and data identifying the associated grammars. Once the machine has been installed on the network N it may communicate automatically with the speech server 2 so that the speech server can determine whether there are new grammars or grammar updates available at step S3 and, if so, can then download these grammars or updates at step S4 in a manner similar to that described above with reference to Figure 12.

[0130] Operation of the system 1c shown in Figure 31 will depend on whether or not the machine to which

instructions ar input (the originating machine) is thesame as the machine which is to carry out the instructions. This is the case for, for example, the copier 9 or 10, the facsimile machine 11 and the multifunction machine 15 when acting as a copier or facsimile machine. In these cases, the operation of the system 1c is as described with reference to Figures 14 to 17. However, when the answer at step S14 in Figure 14 is no it is not possible for the user to be advised of a machine that can carry out this function, because the look-up s rvice 16 is not available. Instead, the steps carried out by the speech operation processor 27 at step S15 in Figure 14 may simply be, as shown in Figure 33, the step S154 of sending to the user a message along the lines of "Function unknown" or "Function unrecognised". Alternatively, when the machine determines that it does not recognise the received control commands, then the machine may send a request to the speech server 2 to s nd text data identifying the unknown command at step S155 in Figure 34 and, once that text data is received at step S156, the sp ech operations processor 27 will cause the main processor 20 of the machine to give the user the message "This machine cannot do ....". For example, where the machine is a copier that cannot do colour copies, then the speech server will send back text data representing the words "colour copies" and the message displayed to the user will be: "This machine cannot do colour copies".

[0131] As a further alternative, text data corresponding to the device control commands may be sent automatically by the speech server 2 so that when the machine determines that it does not recognise the receiv d commands at step S14 it can simply retrieve the accompanying text data and give the appropriate message to the user.

[0132] Figure 35 is a modified form of the flowchart shown in Figure 14 to illustrate the operations of the originating machine when that machine does not itself carry out the requested function (for example where the originating machine is the personal computer or digital camera). As can be seen from Figure 35, in this case the originating machine carries out steps S7 to S9 in the same manner as described above with reference to Figure 14. However, at step S10a, instead of just sending the speech data to the speech server, the originating machine also sends the data to be printed. Figure 36 illustrates the corresponding modifications to the flowchart shown in Figure 15. Communication with the originating machine is carried out at step S19 as described above. At step S20a the speech server 2 receives the speech data together with the data to be printed. Steps S21 to S23 are then carried out as described with reference to Figure 15, and then, at step S24a, the speech server 2 sends the device control commands and the data to be printed to the printer to cause the printer to print iaccordance with the instructions received from the user. Although this arrangement enables instructions to be input to the digital camera 12 or personal computer 13 to control operation of the printer without the use of JINI service objects, it has the disadvantages that there is no possibility of feedback to the user so that if the printer cannot carry out the requested task, it simply produces an error message and that it is necessary for the data to be printed to be transmitted to the speech server and stored at the speech server for subsequent transmission to the printer together with the printer device control commands. This inevitably increases the traffic on the network and the amount of memory storage required by the speech server.

[0133] It will be appreciated that the speech server 2 may also use the information in its machine ID store regarding the other machines on the system (that is the copiers and fax machines) so that the speech server 2 rather than the machine itself can determine whether the machine is capable of carrying out the function requested by the user.

[0134] With the exception of the printing process described with reference to Figures 35 and 36, in the above-described embodiments, a dialogue occurs with the user so that the user is advised where the machine cannot carry out the requested function and is prompted for further instructions when the instructions are incomplete. The present invention may, however, be applied to a system where this facility is not available so that the system responds to a single input instruction by the user and, if the instruction is not understood or the machine cannot carry out the requested function, the user is simply provided with an error message. So that, for example, where the user of the copier 9 issues the instruction "Please copy this", the speech server 2 simply causes the ASR engine 5 to perform speech recognition on this instruction using the copier related grammars from the grammar store 8 and then supplies the appropriate language independent device control commands produced by the interpreter 7 to the copier 9 which, if it understands the received commands, will perform a print operation or otherwise will just issue an error signal. Although such a system should be more simple in operation, it has the disadvantage that the user is provided with no feedback or assistance by the system.

[0135] In each of the embodiments described above, the grammars include rules that determine the structure of valid phrases that can be input as instructions. As an alternative, the grammars may simply include different categories of words that may be expected. For example, in the case of the copy grammar, the copy grammar may include a "action or instruction category" in which words such as copy, make copies, copies, reproduced etc. will be included, a "copy feature" or "manner" section or category in which words such as single-sided, double-sided, reduced, colour etc. will be included and so on. Although such grammars would be more flexible in allowing the user to input instructions in any structure or order, because the structure or order of valid instructions is not constrained, the possibility of misinterpretations and errors will be increased with there being a greater likelihood of confusion between similar words such as "to", "two" and the meaning of the same word in different senses.

[0136] In the embodiments described above, it is assumed that the components additional to the conventional components of the machines 9 to 14 form an integral part of that machine. However, these components may be provided as a separate unit connectable to the main processor 20 of the machine via a linking cable and an appropriate standard interface so that machines such as the copiers 9 and 10, the facsimile machine 11 and the digital camera 12, may be provided both with and without the facility for speech control of their operation. Such a modification may also be implemented in the system shown in Figure 29 although, in this case, the microphone and input speech processing circuitry will in any case already be separate as they are provided by the telephone system.

[0137] In each of the embodiments described above, the speech data may be compressed for transmission to the speech server 2. Generally, the compression algorithmused will be a compression algorithm that compresses the speech data in a manner adapted for the speech recognition engine. Such compression algorithms are known in the art. In the embodiments described above, where use is made of the JAVA/JINI platform then a JINI service object can be used to download to the originating machine the correct speech compression algorithm for the ASR engine to be used for the speech recognition at the speech server.

[0138] In the above described embodiments, the speech server 2 is the only server on the network N. The

speech server may however be dedicated solely to the speech data processing tasks discussed above and a further server or servers may be provided to carry out any other network functions. As discussed above, the network may be any arrangement that enables communication between machines that are physically located in different parts of a building or in different buildings and there need be no network functions to be carried out other than the speech data processing discussed above.

[0139] In the above embodiments, the machines 9 to 15 all form pieces of office equipment. The present invention may, however, be applied to the control of other items of equipment connected over a network such as computer controlled items of machinery and/or computer controlled domestic equipment such as video recorders etc.

[0140] Other modifications will be apparent to those skilled in the art.

Claims

1. A system, comprising:

a first machine couplable to a network and capable of carrying out at least one function;

a second machine couplable to the network;

means for receiving speech data representing instructions spoken by a user and specifying a function to be carried out by the first machine; and

means for transmitting the speech data to the second machine, the second machine having means for accessing speech recognition means for performing speech recognition on received speech data to produce recognised speech data, means for processing recognised speech data to derive from the speech data commands for causing the first machine to carry out the function specified in the spoken instructions and means for transmitting said commands over the network to the first machine, the first machine having means for receiving control commands over the network from the second machine and means responsive to the control commands for causing the function specified by the spoken instructions to be carried out.

2. A system according to claim 1, wherein the speech server has a grammar store which defines words and/or phrases that can be used to in spoken instructions.

3. A system according to claim 1, wherein the speech server has a grammar store storing rules determining the structure of phrases that can be used in spoken input instructions.

4. A system according to claim 1, wherein the speech server has a grammar store storing rules determining the structure of phrases that can be used in spoken instructions and a vocabulary of words that can be used in those phrases.

5. A system according to claim 1, 2, 3 or 4, wherein the first machine is capable of carrying out different functions.

6. A system according to any one of the preceding claims, comprising a plurality of different types of said first machine, each different type being capable of carrying out a different function.

7. A system according to claim 5 or 6 when dependent on any one of claims 2 to 4, wherein the grammar store stores a respective grammar for each different function.

8. A system according to claim 7, wherein the grammar store stores a shared grammar containing words common to instructions relating to the different functions.

9. A network system according to any one of claims 5 to 8, wherein different functions are: photocopying, facsimile transmission and printing.

10. A system according to any one of the preceding claims, wherein speech data receiving means and speech data transmitting means are associated with the or each first machine.

11. A system according to any one of the preceding claims, wherein speech data receiving means and speech data transmitting means are associated with an additional machine coupled to the network.

12. A system according to claim 11, wherein the additional machine is a personal computer or a digital camera and the or at least one of the first machine comprises a printer.

13. A system according to claim 10, 11 or 12, wherein the speech data receiving means and speech data transmitting means form part of the associated machine.

14. A system, comprising:

a speech server couplable to a network;

a plurality of machines couplable to the network, each capable of carrying out at least one function and each having means for receiving control commands over the network and means for carrying out a function in accordance with received control commands;

means for receiving speech data representing instructions spoken by a user and specifying a function to be carried out by a machine; and

means for transmitting th speech data to the speech server;

the speech server having means for accessing sp ech recognition means for performing sp ech recognition on received speech data from on of the machines, means for processing received speech data to derive from the speech data commands for causing the said one machine to carry out the function specified in the spoken instructions and means for transmitting said commands over the network to the said one machine to cause that machine to carry out the specified function.

15. A system according to claim 14, wherein the speech server has a grammar store which defines words and/or phrases that can be used in spoken instructions.

16. A system according to claim 14, wherein the speech server has a grammar store storing rules determining the structure of phrases that can be used in spoken instructions.

17. A system according to claim 14, wherein the speech server has a grammar store storing rules determining the structure of phrases that can used in spoken instructions and a vocabulary of words that can be used in those phrases.

18. A system according to claim 14 to 17, wherein different ones of said plurality of machines are capable of carrying out different functions and the grammar store stores a respective grammar for each different function.

19. A system according to claim 18, wherein the grammar store stores a shared grammar containing words common to instructions relating to the different functions.

20. A system according to any one of claims 14 to 20, wherein the plurality of machines are capable of at least one of photocopying and facsimile transmission functions.

21. A system according to any one of claims 8 to 14, further comprising:

an instruction originating machine coupled to the network having means for receiving speech data representing instructions spoken by a user of the machine and means for transmitting the speech data over the network to the speech server; and

an instruction receiving machine coupled to thenetwork having means for receiving control commands over the network from the speech server and means for causing the instruction receiving machine to act in accordance with the control commands whereby, in use, the instruction receiving machine is caused to carry out the function specified in the instructions spoken by the user of the instruction originating machine.

22. A system, comprising:

a speech server couplable to a network;

an instruction originating machine couplable to the network;

means for receiving speech data representing instructions spoken by a user and specifying a function to be carried out;

means for transmitting speech data to the speech server; and

an instruction receiving machine couplable to the network having means for receiving control commands over the network and means for causing the instruction receiving machine to act in accordance with the control commands,

the speech server having speech recognition means for performing speech recognition on received speech data, means for processing recognised speech data to derive from the speech data commands for causing the instruction receiving machine to carry out the functionspecified by the instructions spoken by the user and means for transmitting said commands over the network to the instruction receiving machine to cause the instruction receiving machine to carry out the function specified by the instructions spoken by the user.

23. A system according to claim 21 or 22, wherein the instruction originating machine is a digital camera or computer and the instruction receiving machine is a printer.

24. A system according to any one of the preceding claims, wherein the speech data receiving means and the speech data transmitting comprise a telephone system coupled to the second machine or server.

25. A system according to claim 24, wherein the telephone system comprises a cordless telephone system.

26. A system according to claim 25, wherein the telephone system comprises a DECT telephone system.

27. A system according to any on of th preceding claims, wherein the speech recognition accessing means is arranged to access a plurality of diff rent speech recognition means and to use the results of speech recognition performed by each of th different speechrecognition means to deriv the recognised speech data.

28. A system according to any one of the preceding claims, wherein the or at least one of the speech recognition means is provided at the second machine or server.

29. A system according to claim 27, wherein at l ast some of the speech recognition means are provided by computers couplable to the network.

30. A system according to any one of the preceding claims, wherein the second machine or server comprises means for identifying the speaker of the instructions and means for accessing speech recognition means trained to the voice of the speaker.

31. A system according to any one of the preceding claims, comprising look-up means for storing information relating to the functions capable of being carried out by machines coupled to the network.

32. A system according to claim 31, further comprising means for determining whether a machine can carry out the function specified in spoken instructions and means for advising the user, on the basis of the information stored in the look-up means, if there is another machine coupledo the network that can carry out that function if the determining means determines that the machine cannot carry out the function.

33. A server for use in a network system, comprising: means for accessing speech recognition means for performing speech recognition to produce recognised speech data on received speech data representing spoken instructions; means for processing recognised speech data to derive from the recognised speech data commands for causing a machine coupled to the network to carry out a function specified by the spoken instructions, and means for transmitting said commands over the network to cause the specified function to be carried out.

34. A server according to claim 33, comprising a grammar store which defines words and/or phrases that can be used in spoken instructions.

35. A server according to claim 33, comprising a grammar store storing rules determining the structure of phrases that can be used in spoken instructions.

36. A server according to claim 33, having a grammar store storing rules determining the structure of phrases that can be in spoken instructions and a vocabulary of words that can be used in those phrases.

37. A server according to any one of claims 34 to 36, wherein the grammar store stores a respective grammar for each of a plurality of different functions that can be carried out by machines coupled to the network.

38. A server according to claim 37, wherein the grammar store stores a shared grammar containing words common to instructions relating to the different functions.

39. A server according to claims 37 or 38, wherein the grammar store stores copy, fax and print grammars.

40. A server according to any one of claims 33 to 39, having means for communicating with a telephone system to receive speech data.

41. A server according to any one of claims 33 to 40, wherein the speech recognition accessing means is arranged to access a plurality of different speech recognition means and to use the results of speech recognition performed by each of the different speech recognition means to derive the recognised speech data.

42. A server according to any one of claims 33 to 41, wherein the or at least one of the speech recognition means is provided at the server.

43. A server according to any one of claims 33 to 42, further comprising means for identifying the speaker of the instructions and means for accessing speech recognition means trained to the voice of the speaker.

44. A machine for carrying out at least one function, comprising:

means for coupling the machine to a network;

means for receiving speech data representing spoken instructions specifying a function to be carried out by the machine;

means for supplying the speech data to speech processing means coupled to the network;

means for receiving from the network control commands derived from the speech data supplied to the speech processing means; and

means responsive to the control commands for causing the function specified by the spoken instructions to be carried out.

45. A photocopying machine comprising:

means for coupling the machine to a network;

means for receiving speech data representing spoken instructions specifying a copy function to be carried out by the machine;

means for supplying the speech data to speech processing means coupled to the network;

means for receiving from the network control commands derived from the sp ech data supplied to the speech processing means; and

means responsive to the control commands for causing the copy function specified by the spoken instructions to be carried out.

46. A facsimile machine comprising:

means for coupling the machine to a network;

means for receiving speech data representing spoken instructions specifying a facsimile function to be carried out by the machine;

means for supplying the speech data to speech processing means coupled to the network;

means for receiving from the network control commands derived from the speech data supplied to the speech processing means; and

means responsive to the control commands for causing the facsimile function specified by the spoken instructions to be carried out.

47. A machine according to claims 44, 45 or 46, further comprising means for determining whether the function specified by the spoken instructions can be carried out and means for advising the user if the function cannot be carried out.

48. A machine according to claim 47, wherein the advising means comprises means for accessing a look-up store containing information relating to the functions that can be carried out by machines coupled to the network and means for advising the user of any other machine that can carry out the requested function.

49. A machine according to claim 45 or 46, wherein the advising means comprise means for causing a message to be displayed on a display of the machine.

50. A digital camera comprising:

means for coupling the camera to a network;

means for receiving speech data representing spoken instructions specifying a print function to be carried out by a printer coupled to the network;

means for supplying the speech data to speech processing means coupled to the network;

means for receiving from the network control commands derived from the speech data supplied to the speech processing means; and

means responsive to control commands received from the network for supplying the control commands together with the data to be printed over the network to the printer for causing the print function specified by the spoken instructions to be carried out by the printer.

51. A digital camera according to claim 50, further comprising means for determining whether the function specified by the spoken instructions can be carried out by the printer and means for advising the user if the function cannot be carried out by the printer.

52. A digital camera according to claim 51, wherein the advising means comprises means for accessing a look-up store containing information relating to the functions that can be carried out by the printers coupled to the network and means for advising the user of any other printers that can carry out the specified function.

53. A device for controlling a machine for carrying out at least one function, comprising:

means for coupling the machine to a network;

means for receiving speech data representing spoken instructions specifying a function to be carried out by the machine;

means for supplying the speech data to speech processing means coupled to the network;

means for receiving from the network control commands derived from the sp ech data supplied to the speech processing means; and

means responsive to th control commands for supplying to the machine commands for causing the function specified by the spoken instructions to b carried out.

54. A device for controlling a photocopying machine comprising:

means for coupling the machine to a network;

means for receiving speech data representing spoken instructions specifying a copy function to be carried out by the machine;

means for supplying the speech data to speech processing means coupled to the network;

means for receiving from the network control commands derived from the speech data supplied to the speech processing means; and

means responsive to the control commands for supplying to the machine commands for causing the copy function specified by the spoken instructions to be carried out.

55. A device for controlling a facsimile machine comprising:

means for coupling the machine to a network;

means for receiving speech data representing spoken instructions specifying a facsimile function to be carried out by the machine;

means for supplying the speech data to speech processing means coupled to the network;

means for receiving from the network controlcommands derived from the speech data supplied to the speech processing means; and

means responsive to the control commands for supplying to the machine commands for causing the facsimile function specified by the spoken instructions to be carried out.

56. A device for controlling a digital camera, comprising:

means for coupling the camera to a network;

means for receiving speech data representing spoken instructions specifying a print function to be carried out by a printer coupled to the network;

means for supplying the speech data to speech processing means coupled to the network;

means for receiving from the network control commands derived from the speech data supplied to the speech processing means; and

means responsive to control commands received from the network for supplying the control commands together with the data to be printed over the network to the printer for causing the print function specified by the spoken instructions to be carried out by the printer.

57. A device according to claims 53, 54, 55 or 56, further comprising means for determining whether the function specified by the spoken instructions can becarried out and means for advising the user if the function cannot be carried out.

58. A device according to claim 57, wherein the advising means comprises means for accessing a look-up store containing information relating to the functions that can be carried out by machines coupled to the network and means for advising the user of any other machine that can carry out the requested function.

59. A device according to claim 57 or 58, wherein the advising means comprise means for causing a message to be displayed on a display of the machine.

60. A method of controlling operation of a machine coupled to a network, comprising:

receiving speech data representing spoken instructions for controlling the machine;

performing speech recognition on the received speech data to produce recognised speech data;

processing the recognised speech data to derive from the recognised speech data commands for causing

th machine to carry out a function specified by the spoken instructions; and

transmitting said commands over the network to cause the specified function to b carried out.

61. A method according to claim 60, which comprises using a grammar store which defines words and/or phrases that can be used in spoken instructions to perform said speech recognition.

62. A method according to claim 60, which comprises using a grammar store storing rules determining the structure of phrases that can be used in spoken instructions to perform said speech recognition.

63. A method according to claim 60, which comprises using a grammar store storing rules determining the structure of phrases that can be in spoken instructions and a vocabulary of words that can be used in those phrases to perform said speech recognition.

64. A method according to any one of claims 60 to 63, which comprises using as said grammar store a grammar store storing a respective grammar for each of a plurality of different functions that can be carried out by machines coupled to the network.

65. A method according to claim 64, which comprises using as said grammar store a grammar store storing a shared grammar containing words common to instructions relating to the different functions.

66. A method according to claim 64 or 65, which comprises using a grammar store storing copy, fax and print grammars.

67. A method according to any one of claims 60 to 66, which comprises receiving speech data over a telephone system.

68. A method according to any one of claims 60 to 67, which comprises performing speech recognition using a plurality of different speech recognition means and deriving the recognised speech data using the results of speech recognition performed by each of the different speech recognition means.

69. A method according to any one of claims 60 to 69, further comprising identifying the speaker of the instructions and accessing speech recognition means trained to the voice of the speaker to perform the spe ch recognition.

70. A method of operating a machine, comprising:

receiving speech data representing spoken instructions specifying a function to be carried out by the machine;

supplying the speech data to a network;

receiving from the network control commands derivedfrom the speech data supplied to the network; and

causing the function specified by the spoken instruction to be carried out in response to receipt of the control commands.

71. A method of operating a photocopying machine, comprising:

receiving speech data representing spoken instructions specifying a copy function to be carried out by the machine;

supplying the speech data to a network;

receiving from the network control commands derived from the speech data supplied to the network; and

causing the copy function specified by the spoken instructions to be carried out in response to receipt of the control commands.

72. A method of operating a facsimile machine, comprising:

receiving speech data representing spoken instructions specifying a facsimile function to be carried out by the machine;

supplying the speech data to a network;

receiving from the network control commands derived from the speech data supplied to the network; and

causing the facsimile function specified by the spoken instructions to be carried out in response toreceipt of the control commands.

73. A method according to claim 70, 71 or 72, further comprising determining whether the function specified by the spoken instructions can be carried out and advising the user if the function cannot be carried out.

74. A method according to claim 73, which comprises advising the user by accessing a look-up store containing information relating to the functions that can be carried out by machines coupled to the network and then advising the user of any other machine that can carry out the requested function.

75. A method according to claim 73 or 74, which comprises advising the user by causing a message to be displayed on a display of the machine.

76. A method of operating a digital camera, comprising:

receiving speech data representing spoken instructions specifying a print function to be carried out by a printer coupled to a network;

supplying the speech data to the network;

receiving from the network control commands derived from the speech data supplied to the network; and, in response to receipt of the control commands, supplying the control commands together with the data to be printed over the network to the printer for causing the print function specified by the spoken instructions to be carried out by the printer.

77. A method according to claim 76, further comprising determining whether the function specified by the spoken instructions can be carried out by the printer and advising the user if the function cannot be carried out by the printer.

78. A method according to claim 77, which comprises advising the user by accessing a look-up store containing information relating to the functions that can be carried out by the printers coupled to the network and then advising the user of any other printers that can carry out the specified function.

79. A system, comprising:

a plurality of first machines of different types couplable to a network;

a second machine couplable to the network;

means for receiving speech data representing instructions spoken by a user; and means for transmitting the speech data to the second machine,

the second machine having means for accessing speech recognition means for performing speech recognition on received speech data to produce recognised speech data,means for processing recognised speech data to derive from the speech data commands for supply to a first machine and means for transmitting said commands over the network to that first machine, the accessing means being arranged to access a shared grammar or set of grammars for at least one of machines of the same type and of functions of the same type.

80. A system, comprising:

a plurality of first machines of different types couplable to a network;

a second machine couplable to the network;

means for receiving speech data representing instructions spoken by a user; and means for transmitting the speech data to the second machine;

the second machine having means for accessing speech recognition means for performing speech recognition on received speech data to produce recognised speech data, means for processing recognised speech data to derive from the speech data commands for supply to a first machine and means for transmitting said commands over the network to that first machine, the accessing means being arranged to access a different grammar or a different set of grammars for each different type of first machine.

81. A system, comprising:

a plurality of first machines couplable to anetwork;

a second machine couplable to the network;

means for receiving speech data representing instructions spoken by a user; and means for transmitting the speech data to the second machine,

the second machine having means for accessing speech recognition means for performing speech recognition on received speech data to produce recognised speech data, means for processing recognised

speech data to derive from the sp ech data commands for supply to a first machine and means for transmitting said commands over the network to that first machine, the accessing m ans being arranged to access a grammar store associating each of a number of voice macros or spoken words or phrases with a series of functions to be carried out at a first machine.

82. A system, comprising:

at least one first machine couplabl to a network;

a second machine couplable to the network;

means for receiving speech data representing instructions spoken by a user; and

means for transmitting the speech data to the second machine, the second machine having means for accessing speech recognition means for performing speech recognition on received speech data to produce recognised speech data, means for processing recognised speech datato derive from the speech data commands for supply to the or a first machine and means for transmitting said commands over the network to that first machine, that first machine having means for receiving control commands over the network from the second machine, means for determining whether the first machine can process the received commands and means for advising the user of another machine on the network that can process the rec ived commands if the determining means determines that that first machine cannot.

83. A system, comprising:

at least one first machine couplable to a network;

a second machine couplable to the network;

means for receiving speech data representing instructions spoken by a user; and

means for transmitting the speech data to the second machine, the second machine having means for accessing a plurality of different speech recognition means for performing speech recognition on received speech data, means for comparing the results of speech recognition carried out by the different speech recognition means, means for using the comparison of the results to produce recognised speech data, means for processing recognised speech data to derive from the speech data commands for supply to the or a first machine and means for transmitting said commands over the network to that firstmachine.

84. A system according to any one of claims 79 to 83, wherein the speech data receiving means and the speech data transmitting means are provided at the or a first machine and the speech data transmitting means is arranged to transmit speech data over the network.

85. A system, comprising:

at least one first machine couplable to a network;

a second machine couplable to the network;

a telephone system for receiving speech data representing instructions spoken by a user and for transmitting the speech data to the second machine,

the second machine having means for accessing speech recognition means for performing speech recognition on received speech data to produce recognised speech data, means for processing recognised speech data to derive from the speech data commands for supply to the or a first machine and means for transmitting said commands over the network to the first machine.

86. A system according to any one of claims 1 to 32 or 79 to 85, comprising a shared grammar which defines words that can be used in spoken instructions in relation to any machine of a particular type, for example photocopiers.

87. A system according to claim 86, comprising means for determining if a machine can carry out an identified function using the shared grammar and means for advising the user if the machine is not capable of carrying out the identified function.

88. A system according to any one of claims 1 to 30 or 79 to 85, comprising a look up service containing information regarding the machines coupled to the network, means for determining whether there is any machine coupled to the network that can carry out the identified function and means for advising the user of the results of the determination by the determining means.

89. A system according to any one of claims 1 to 32 or 79 to 88, comprising a start up grammar, means for identifying a machine type on the basis of speech recognized using the start up grammar, and means for selecting a grammar which defines words that can be used in spoken instructions in relation to the identified machine type.

90. A system according to any one of claims 1 to 32 or 79 to 89, wherein a machine comprises a manually operable user interface for receiving manually input commands for causing the machine to carry out a function, the machine being operable to carry out a function in accordance with manually input commands, spoken commands or a combination of manually input and spoken commands.

91. A system according to any one of claims 1 to 32 or 79 to 90, comprising means for compressing speech data for transmission by the transmitting means.

92. A system, comprising:

a plurality of machines capable of carrying out at least one function and each comprising a Java virtual machine connectable to a network;

speech data processing means comprising a Java virtual machine connectable to the network, the speech data processing means being operable to receive speech data input by a user and to generate from that speech data control commands for transmission over the network to cause a machine to carry out the function specified in the received speech data;

a JINI look up service couplable to the network;

means for registering a machine with the JINI look up service when that machine is coupled to the network, the speech data processing means being operable to access the JINI look up service to determine whether a machine is capable of carrying out a function specified in received speech data.

93. A system comprising:

an originating machine comprising a Java virtual machine connectable to a network;

at least one printer comprising a Java virtual machine connectable to the network;

speech data processing means comprising a Java virtual machine connectable to the network, the speech data processing means being operable to receive from a user of the originating machine speech data representing a request for printing by the printer of data to be provided over the network by the originating machine, the speech data processing means being operable to download a JINI service object associated with the printer to the originating machine in response to receipt of such a request and to send to the JINI service object commands for causing the printer to carry out a printing function in response to speech data received from the user, the JINI service object being operable to determine whether the printer can carry out the requested printing function and, if so, to communicate with the printer over the network to cause the printer to carry out the requested printing function.

94. A system comprising:

an originating machine comprising a Java virtual machine connectable to a network;

at least one printer comprising a Java virtual machine connectable to the network;

a speech data processing means comprising a Java virtual machine connectable to the network, the speech data processing means being operable to receive from a user of the originating machine speech data representing a request for printing by the printer of data to be provided over the network by the originating machine;

a JINI look up service containing a directory of printers comprising Java virtual machines connected to the network and identifying the functions that can be carried out by those printers;

the speech data processing means being operable to determine from speech data input by the user the printing function required by the user, to select a compatible printer using the JINI look up service, and to download a JINI service object associated with the compatible printer to the originating machine in response to receipt of such a request, the speech data processing means being operable to send to the JINI service object commands for causing the printer to carry out a printing function in response to speech data received from the user representing those commands, and the JINI service object being operable to determine whether the printer can carry out the requested printing function and, if so, to communicate with the printer over the network to cause the printer to carry out the requested printing function.

95. A system comprising:

a plurality of machines each capable of carrying out at least one function and each comprising a Java virtual machine connectable to a network;

a JINI look up service containing a directory of Java virtual machines connected to the network and identifying the functions that can be carried out by those machines; and

speech data processing means comprising a Java virtual machine connectable to the network, the speech

data processing means being operable to receive speech data input by a user, to identify a function represented by received speech data and to access the JINI look up service to identify any machine coupled to the network that is capable of carrying out that function.

96. A system according to any one of claims 92 to 95, comprising a grammar specific to the JINX look up service accessible by the speech data processing means .

97. A system, comprising:

a plurality of machines capable of carrying out at least one function and each comprising a Java virtual machine connectable to a network;

speech data processing means comprising a Java virtual machine connectable to the network, the speech data processing means being operable to receive speech data input by a user and to generate from that speech data control commands for transmission over the network to cause a machine to carry out the function specified in the received speech data; and

means for using a JINI service object to download a compression algorithm for compressing speech data for transmission.

98. A system according to any one of claims 1 to 32 or 79 to 97, wherein the speech data processing means is operable to receive speech data input by the user using a DECT telephone and means are provided for communicating with the user via the user's DECT telephone.

99. A system according to claim 98, comprising means for identifying a machine at which a user is located in accordance with the identity of the user's DECT telephone.

100. A system according to claim 99, comprising means for selecting a speech recognition engine to be used by the speech data processing means in accordance with the identity of the DECT telephone.

101. A system according to any one of claims 98 to 100, comprising means for enabling a user to defer carrying out of a function by a machine, the system being arranged to transmit commands over the network to the machine to cause the machine to carry out the function in response to speech data received via the user's DECT telephone.

102. A machine for use in a network as claimed in any one of claims 79 to 101 having the machine features set out in any one or any combination of claims 79 to 101.

103. A signal carrying processor implementable instructions for causing processing means to be configured to provide a network system as claimed in any one of claims 1 to 31 and 79 to 101, a server as claimed in any one of claims 32 to 43, a machine as claimed in any one of claims 44 to 49, 86 or 87 or a digital camera as claimed in any one of claims 50 to 52 or a device as claimed in any one of claims 53 to 59.

104. A storage medium carrying processor implementable instructions for causing processing means to be configured to provide a network system as claimed in any one of claims 1 to 31 and 79 to 101, a server as claimed in any one of claims 32 to 43, a machine as claimed in any one of claims 44 to 49, 86 or 87 or a digital camera as claimed in any one of claims 50 to 52 or a device as claimed in any one of claims 53 to 59.

105. A signal carrying processor implementable instructions for causing processing means to carry out a method as claimed in any one of claims 60 to 78.

106. A storage medium carrying processor implementable instructions for causing processing means to carry out a method as claimed in any one of claims 60 to 78.

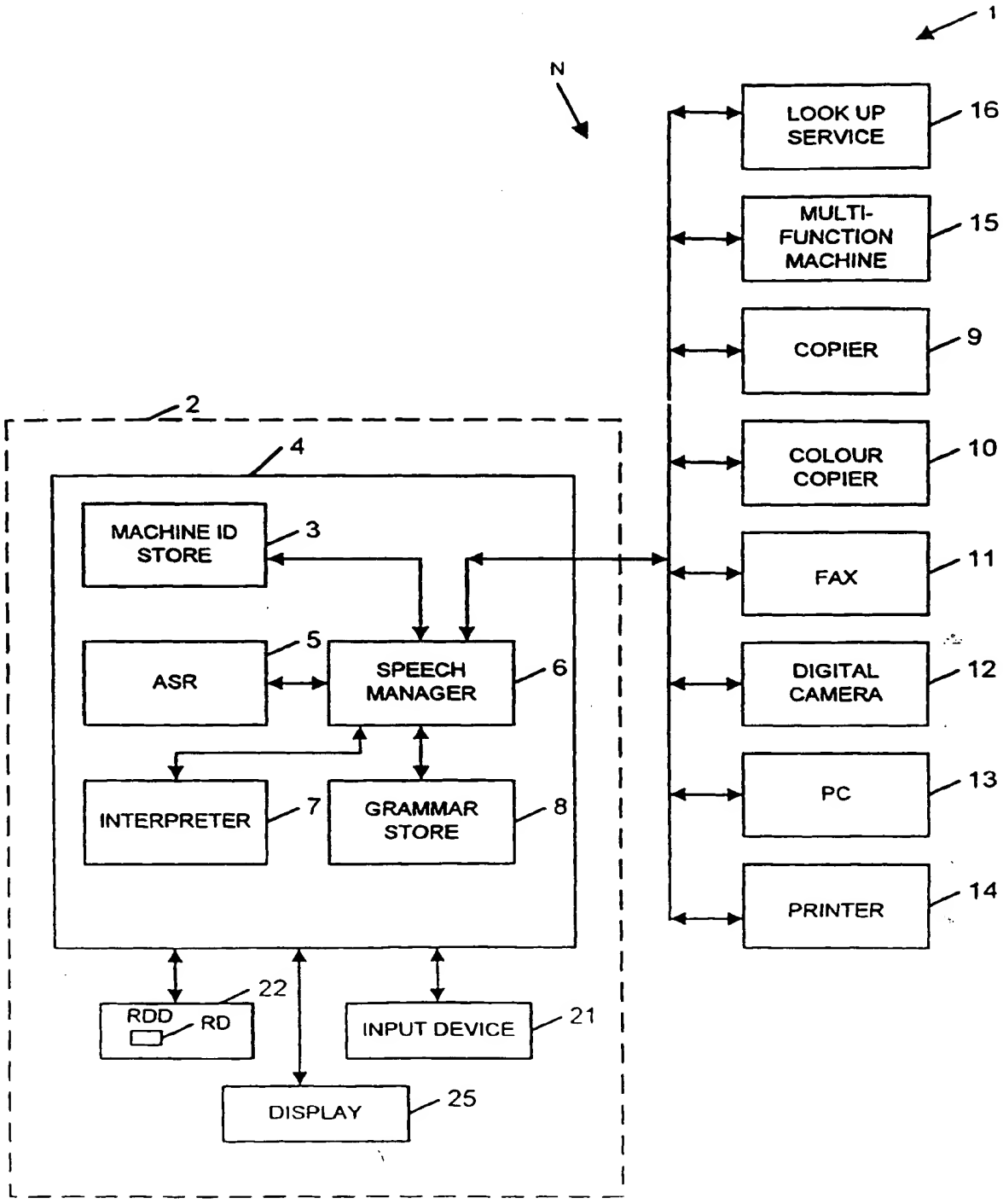

1
N
LOOK UP SERVICE
16
MULTI-FUNCTION MACHINE
15
COPIER
9
COLOUR COPIER
10
FAX
11
DIGITAL CAMERA
12
PC
13
PRINTER
14
2
4
MACHINE ID STORE
3
ASR
5
SPEECH MANAGER
6
INTERPRETER
7
GRAMMAR STORE
8
RDD
RD
22
INPUT DEVICE
21
DISPLAY
25

FIG. 1

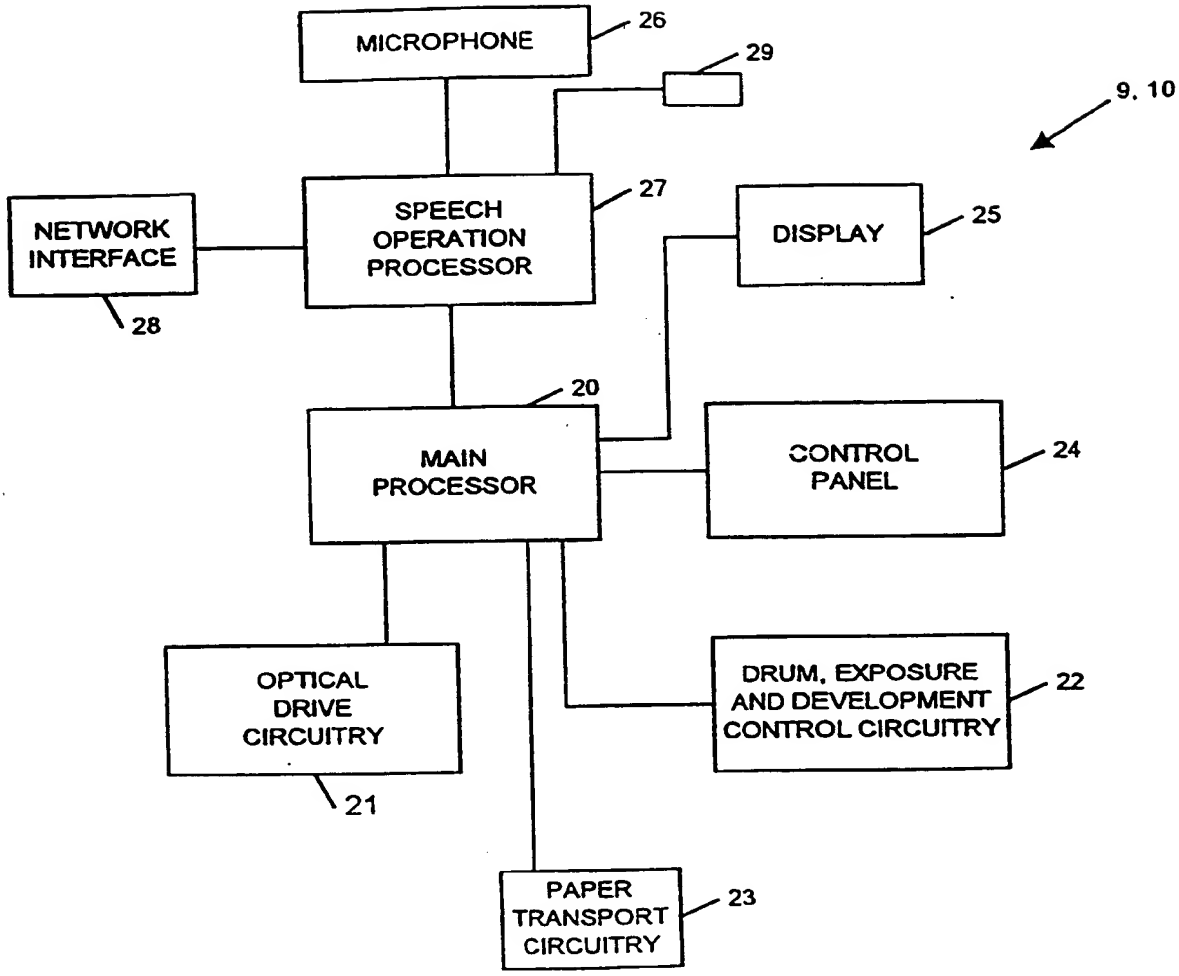

MICROPHONE
26
29
9, 10
NETWORK INTERFACE
28
SPEECH OPERATION PROCESSOR
27
DISPLAY
25
20
MAIN PROCESSOR
CONTROL PANEL
24
OPTICAL DRIVE CIRCUITRY
21
DRUM, EXPOSURE AND DEVELOPMENT CONTROL CIRCUITRY
22
PAPER TRANSPORT CIRCUITRY
23

FIG. 2

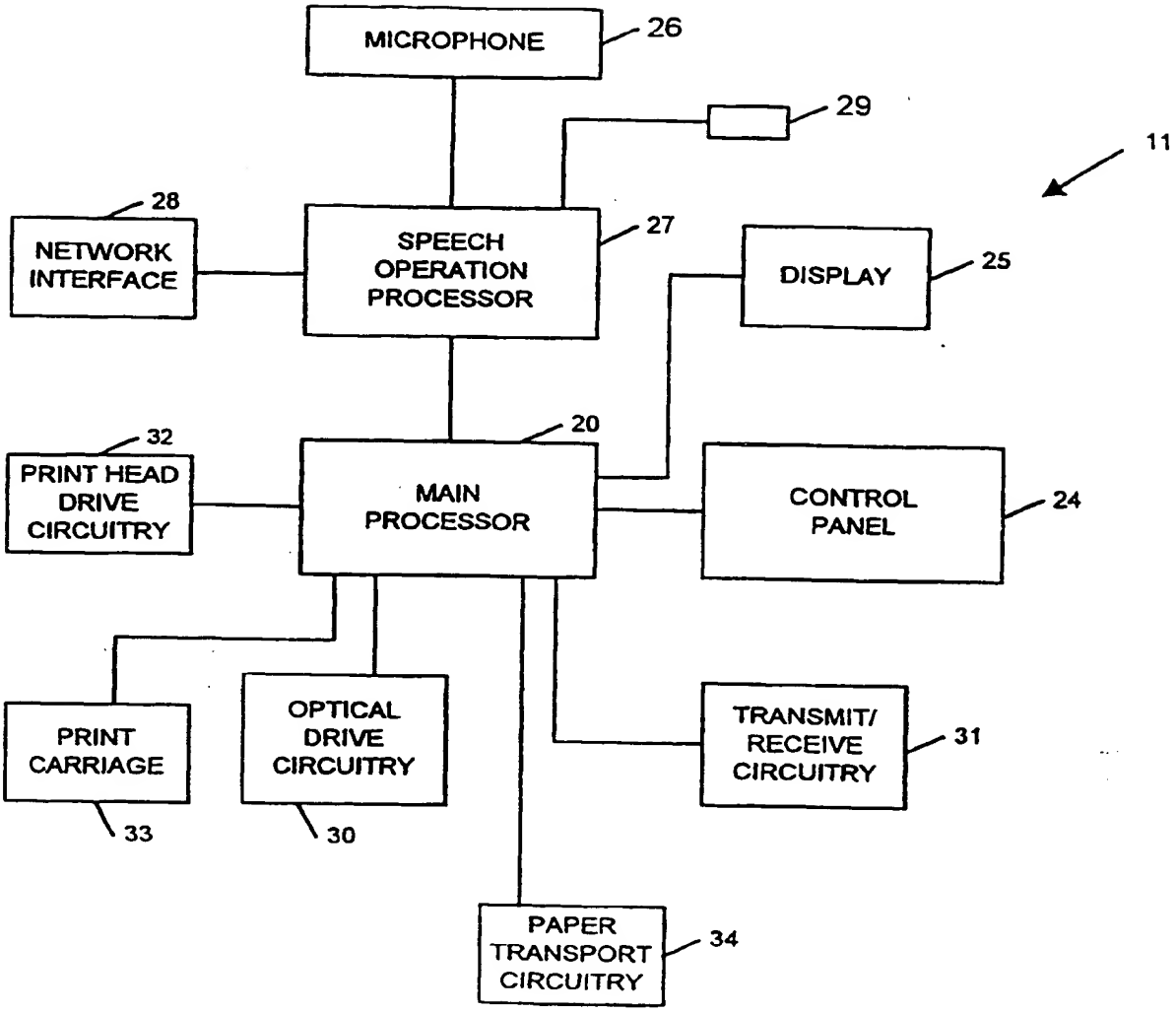

MICROPHONE
26
29
11
28
NETWORK INTERFACE
SPEECH OPERATION PROCESSOR
27
DISPLAY
25
32
PRINT HEAD DRIVE CIRCUITRY
20
MAIN PROCESSOR
CONTROL PANEL
24
PRINT CARRIAGE
33
OPTICAL DRIVE CIRCUITRY
30
TRANSMIT/ RECEIVE CIRCUITRY
31
PAPER TRANSPORT CIRCUITRY
34

FIG. 3

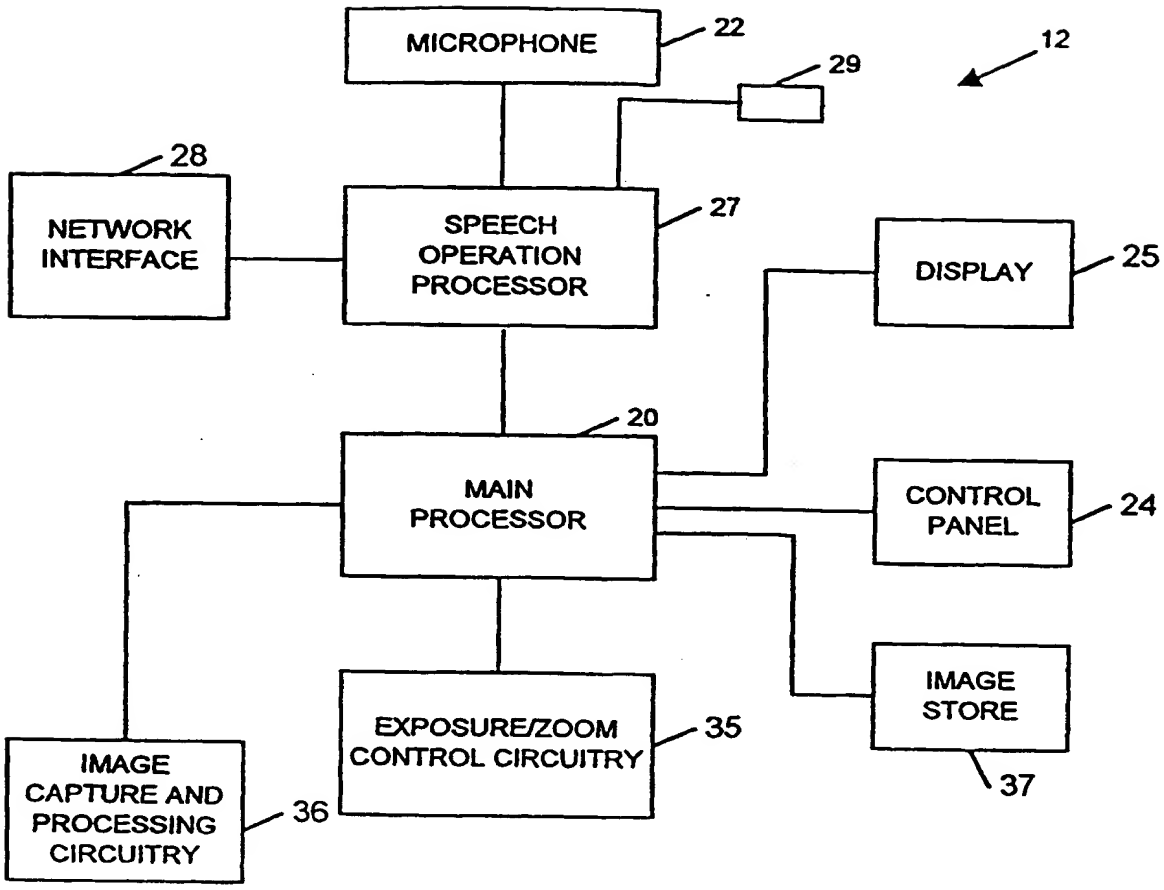

MICROPHONE
22
29
12
28
NETWORK INTERFACE
SPEECH OPERATION PROCESSOR
27
DISPLAY
25
20
MAIN PROCESSOR
CONTROL PANEL
24
IMAGE STORE
37
EXPOSURE/ZOOM CONTROL CIRCUITRY
35
IMAGE CAPTURE AND PROCESSING CIRCUITRY
36

FIG. 4

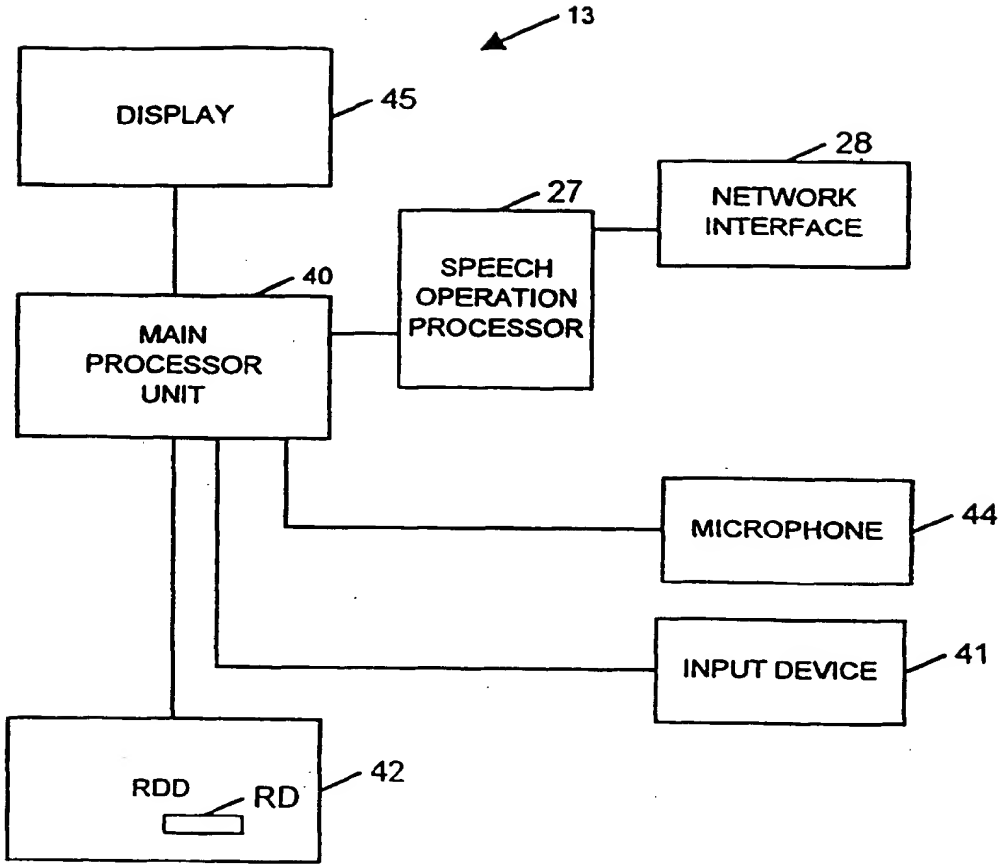

13
DISPLAY
45
28
NETWORK INTERFACE
27
SPEECH OPERATION PROCESSOR
40
MAIN PROCESSOR UNIT
MICROPHONE
44
INPUT DEVICE
41
RDD
RD
42

FIG. 5

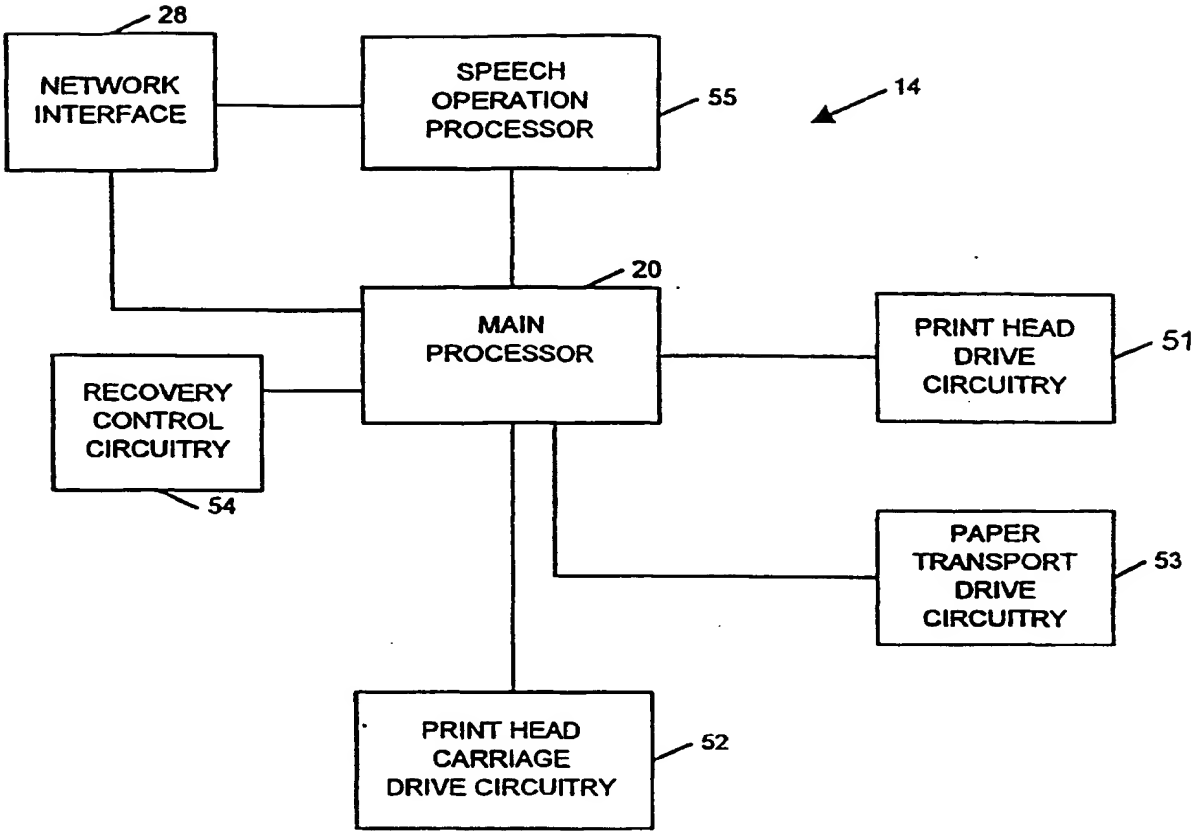

28
NETWORK INTERFACE
SPEECH OPERATION PROCESSOR
55
14
20
MAIN PROCESSOR
PRINT HEAD DRIVE CIRCUITRY
51
RECOVERY CONTROL CIRCUITRY
54
PAPER TRANSPORT DRIVE CIRCUITRY
53
PRINT HEAD CARRIAGE DRIVE CIRCUITRY
52

FIG. 6

COPY GRAMMAR	
COPY	SINGLE-SIDED
MAKE COPY	DOUBLE-SIDED
MAKE COPIES	REDUCED
COPIES	COLOUR
REPRODUCE	BLACK & WHITE
	COLLATED
	A4
	A3
	STAPLED

FIG. 7

PRINT GRAMMAR		
PRINT	COLOUR	LASER
HARD COPY	HIGH RESOLUTION	COLOUR LASER
REPRODUCE	PHOTO-QUALITY	INK JET
	REDUCED	COLOUR INK JET
	BLACK & WHITE	
	SAME SIZE	

FIG. 8

FAX GRAMMAR	
FAX	STORED NAMES
SEND	SHORT CODES
	COUNTRY CODES
	AREA CODES

FIG. 9

SHARED WORDS GRAMMAR		
YES	BY	PAPER
THAT	ONCE	PAGE
THIS	TWICE	TEXT
PLEASE	NEXT	PICTURE
COULD YOU	PAST	IMAGE
AND	DOCUMENT	NO

FIG. 10

TIME/DATE GRAMMAR	
AM	TODAY
PM	TOMORROW
O'CLOCK	MONDAY....SUNDAY
MORNING	JANUARY...DECEMBER
AFTERNOON	
EVENING	

FIG. 11

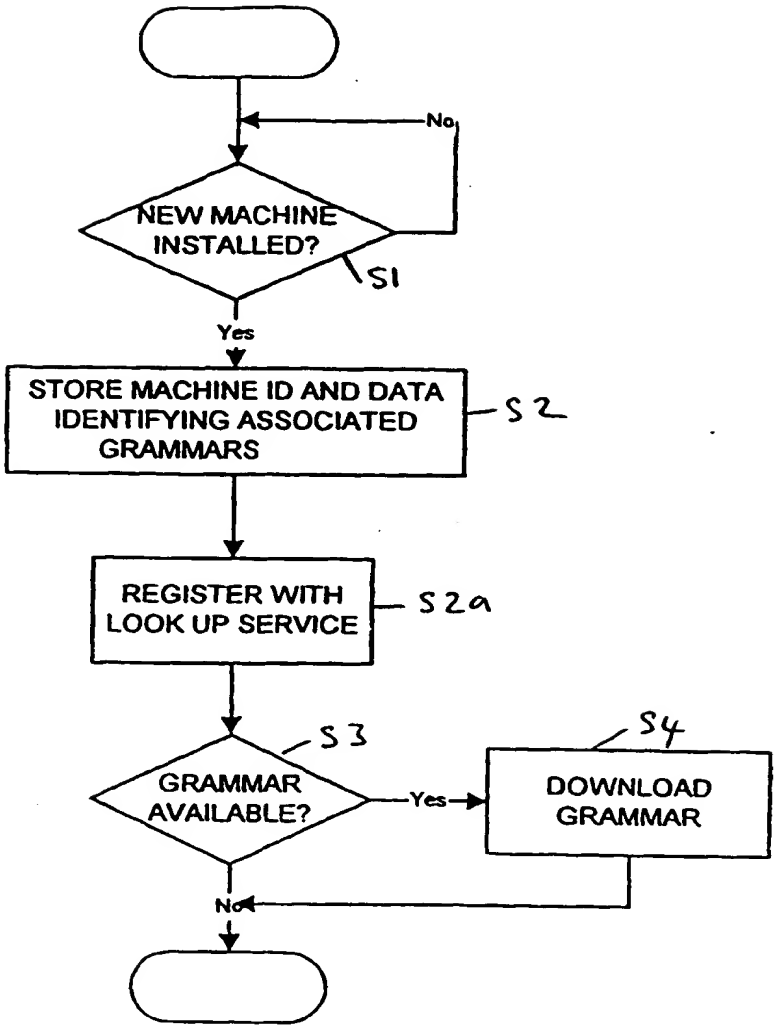

No
NEW MACHINE INSTALLED?
S1
Yes
STORE MACHINE ID AND DATA IDENTIFYING ASSOCIATED GRAMMARS
S2
REGISTER WITH LOOK UP SERVICE
S2a
S3
GRAMMAR AVAILABLE?
Yes
S4
DOWNLOAD GRAMMAR
No

FIG. 12

MACHINE ID	COPIER NO 9
ASSOCIATED GRAMMARS	COPY; SHARED WORDS; TIME/DATE
FUNCTIONS	
BLACK & WHITE	YES
COLOUR	NO
DOUBLE-SIDED	YES
REDUCED	YES
COLLATED	NO
A4	YES
A3	NO
STAPLED	NO

FIG. 13

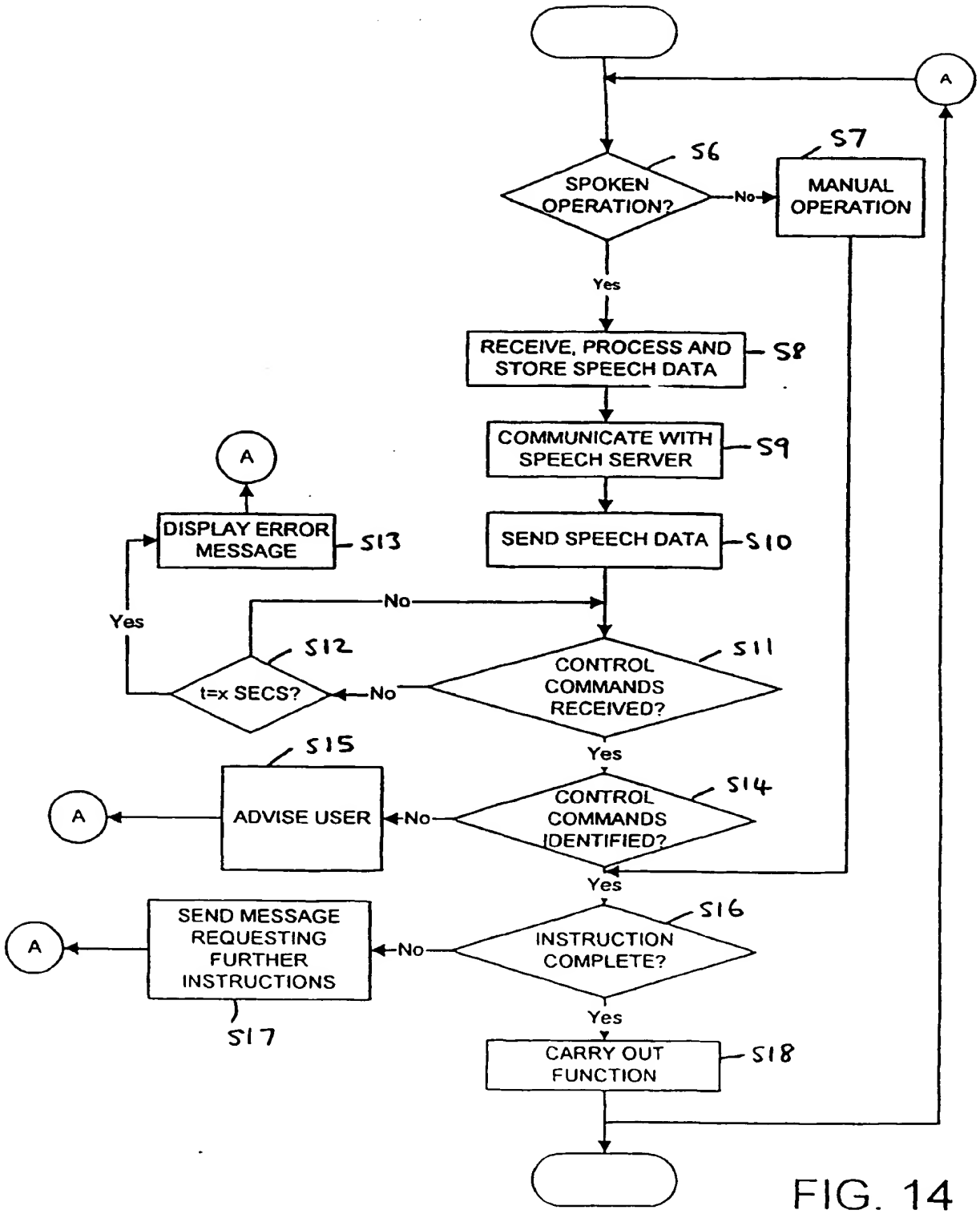

A
S6
SPOKEN OPERATION?
No
S7
MANUAL OPERATION
Yes
RECEIVE, PROCESS AND STORE SPEECH DATA
S8
COMMUNICATE WITH SPEECH SERVER
S9
SEND SPEECH DATA
S10
CONTROL COMMANDS RECEIVED?
S11
No
t=x SECS?
S12
No
Yes
DISPLAY ERROR MESSAGE
S13
A
Yes
CONTROL COMMANDS IDENTIFIED?
S14
No
ADVISE USER
S15
A
Yes
INSTRUCTION COMPLETE?
S16
No
SEND MESSAGE REQUESTING FURTHER INSTRUCTIONS
S17
A
Yes
CARRY OUT FUNCTION
S18

FIG. 14

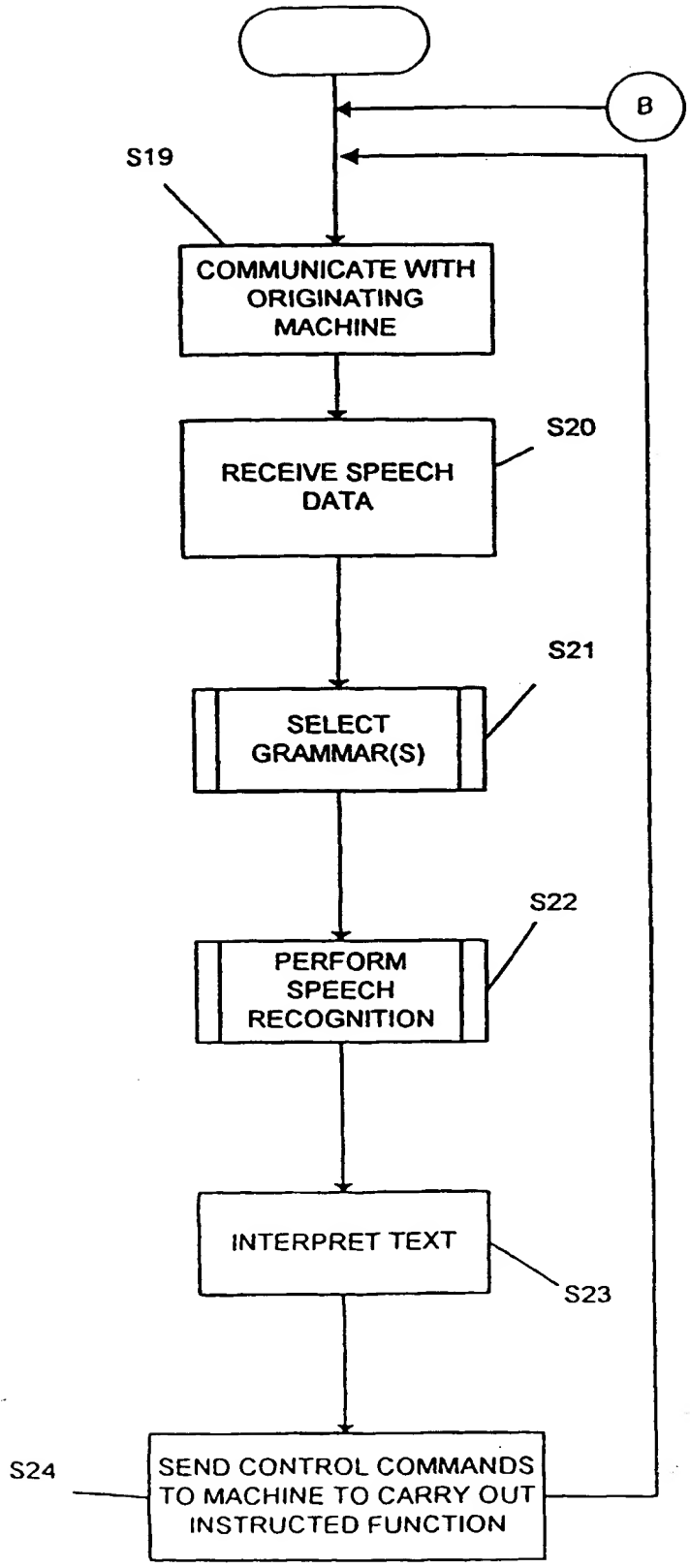

B
S19
COMMUNICATE WITH ORIGINATING MACHINE
S20
RECEIVE SPEECH DATA
S21
SELECT GRAMMAR(S)
S22
PERFORM SPEECH RECOGNITION
INTERPRET TEXT
S23
S24
SEND CONTROL COMMANDS TO MACHINE TO CARRY OUT INSTRUCTED FUNCTION

FIG. 15

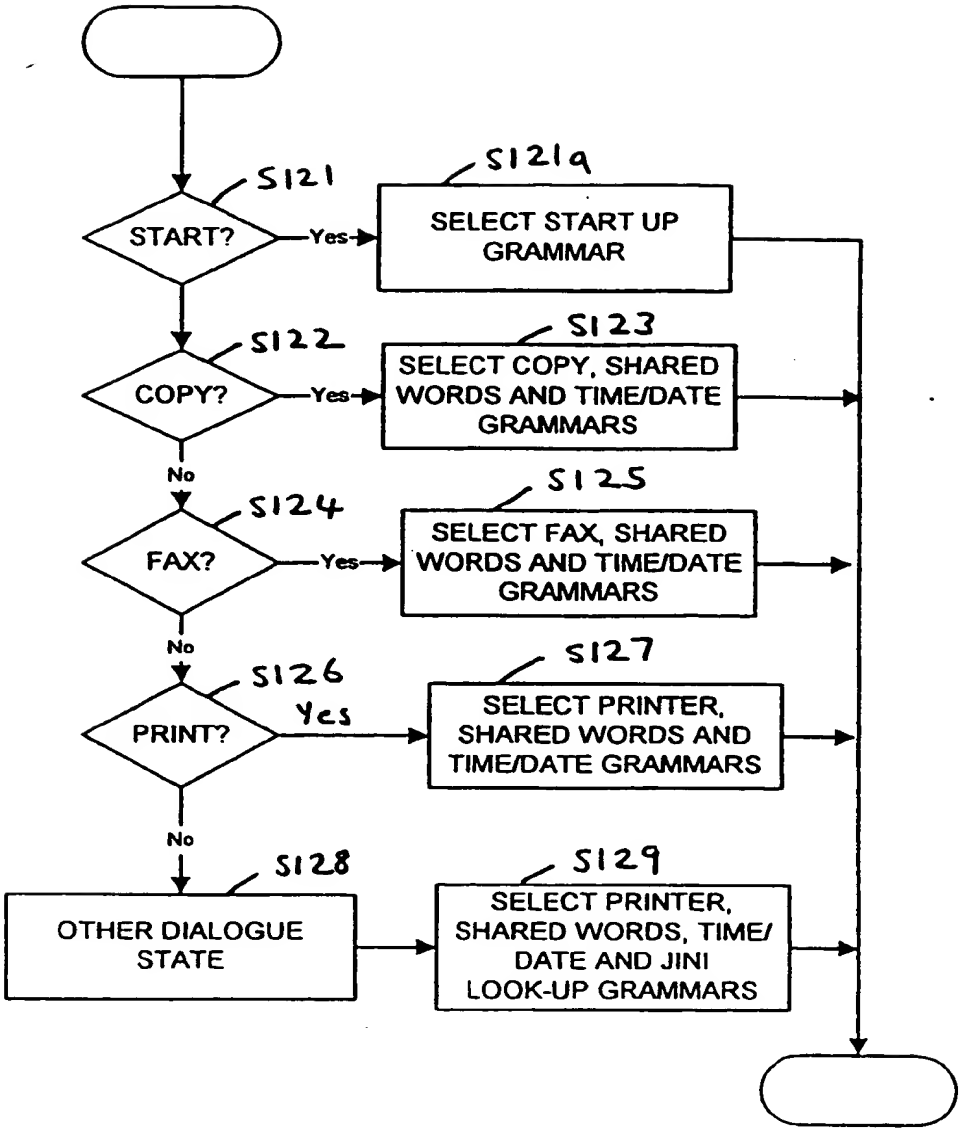

S121
START?
Yes
S121a
SELECT START UP GRAMMAR
S122
COPY?
Yes
S123
SELECT COPY, SHARED WORDS AND TIME/DATE GRAMMARS
No
S124
FAX?
Yes
S125
SELECT FAX, SHARED WORDS AND TIME/DATE GRAMMARS
No
S126
PRINT?
Yes
S127
SELECT PRINTER, SHARED WORDS AND TIME/DATE GRAMMARS
No
S128
OTHER DIALOGUE STATE
S129
SELECT PRINTER, SHARED WORDS, TIME/ DATE AND JINI LOOK-UP GRAMMARS

FIG. 16

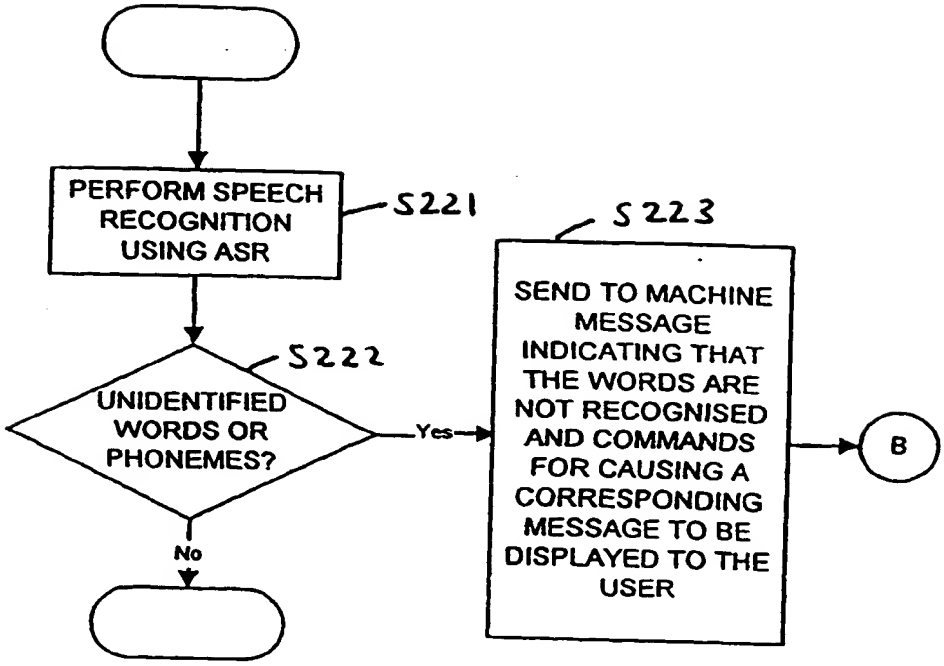

PERFORM SPEECH RECOGNITION USING ASR
S221
S222
UNIDENTIFIED WORDS OR PHONEMES?
Yes
No
S223
SEND TO MACHINE MESSAGE INDICATING THAT THE WORDS ARE NOT RECOGNISED AND COMMANDS FOR CAUSING A CORRESPONDING MESSAGE TO BE DISPLAYED TO THE USER
B

FIG. 17

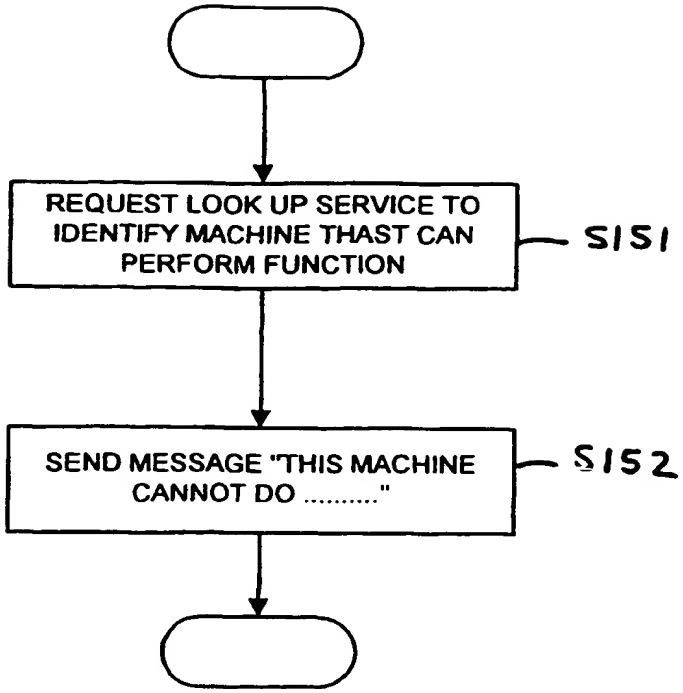

REQUEST LOOK UP SERVICE TO IDENTIFY MACHINE THAST CAN PERFORM FUNCTION
S151
SEND MESSAGE "THIS MACHINE CANNOT DO ....."
S152

FIG. 18

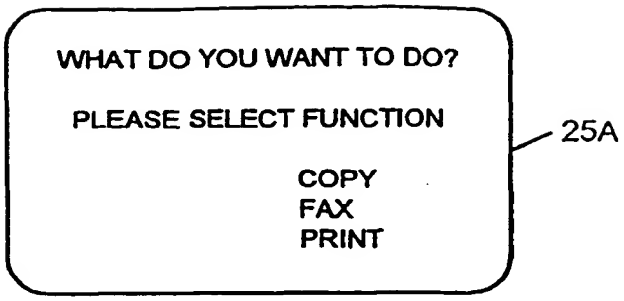


FIG. 19

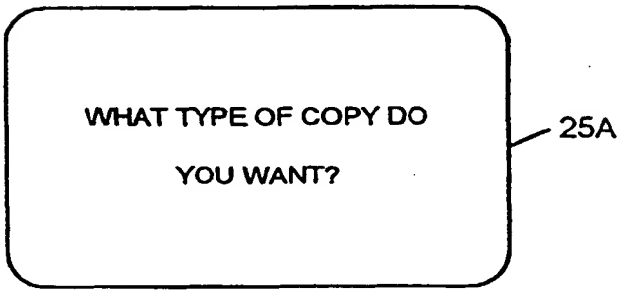


FIG. 20

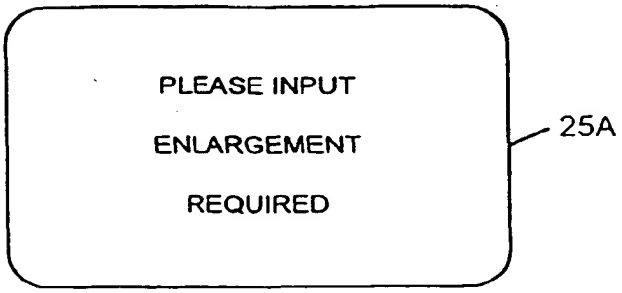


FIG. 21

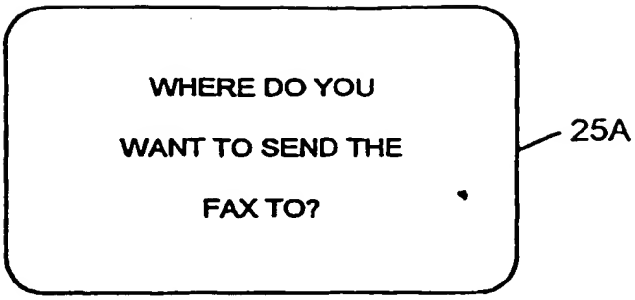


FIG. 22

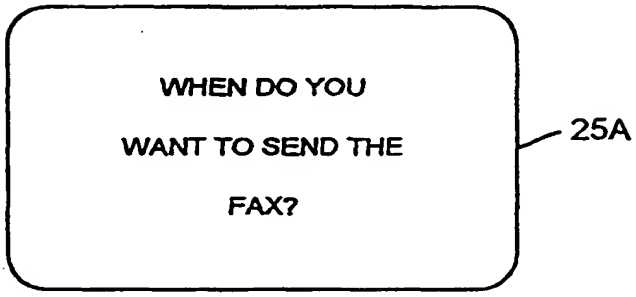


FIG. 23

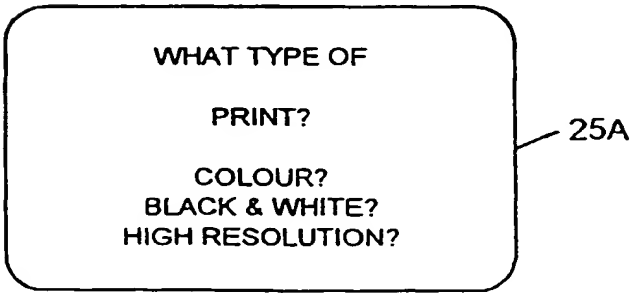


FIG. 24

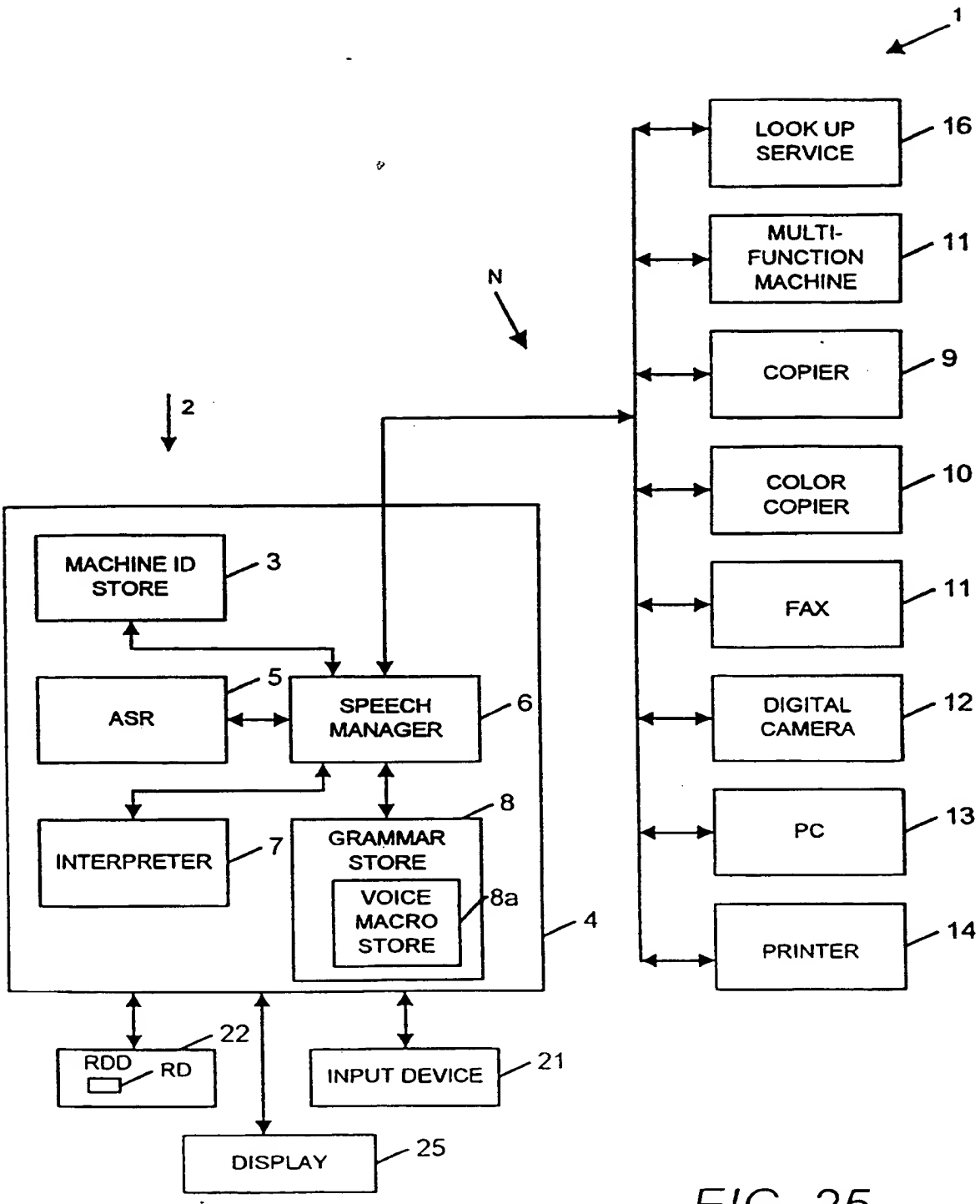

1
LOOK UP SERVICE
16
MULTI-FUNCTION MACHINE
11
N
COPIER
9
2
COLOR COPIER
10
MACHINE ID STORE
3
FAX
11
ASR
5
SPEECH MANAGER
6
DIGITAL CAMERA
12
PC
13
INTERPRETER
7
GRAMMAR STORE
8
VOICE MACRO STORE
8a
4
PRINTER
14
RDD
RD
22
INPUT DEVICE
21
DISPLAY
25

FIG. 25

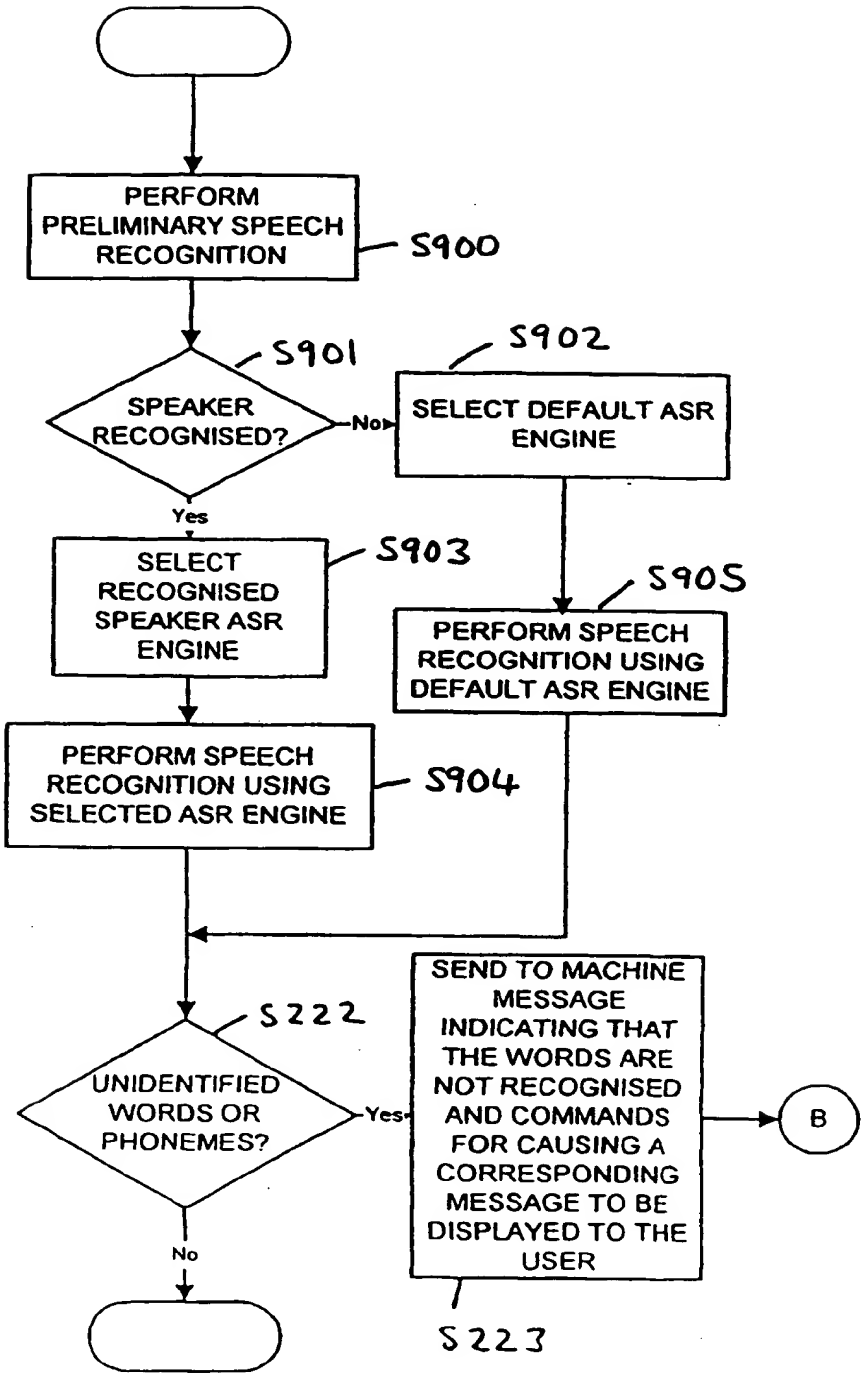

PERFORM PRELIMINARY SPEECH RECOGNITION
S900
S901
SPEAKER RECOGNISED?
No
S902
SELECT DEFAULT ASR ENGINE
Yes
SELECT RECOGNISED SPEAKER ASR ENGINE
S903
S905
PERFORM SPEECH RECOGNITION USING DEFAULT ASR ENGINE
PERFORM SPEECH RECOGNITION USING SELECTED ASR ENGINE
S904
S222
UNIDENTIFIED WORDS OR PHONEMES?
Yes
SEND TO MACHINE MESSAGE INDICATING THAT THE WORDS ARE NOT RECOGNISED AND COMMANDS FOR CAUSING A CORRESPONDING MESSAGE TO BE DISPLAYED TO THE USER
B
No
S223

FIG. 26

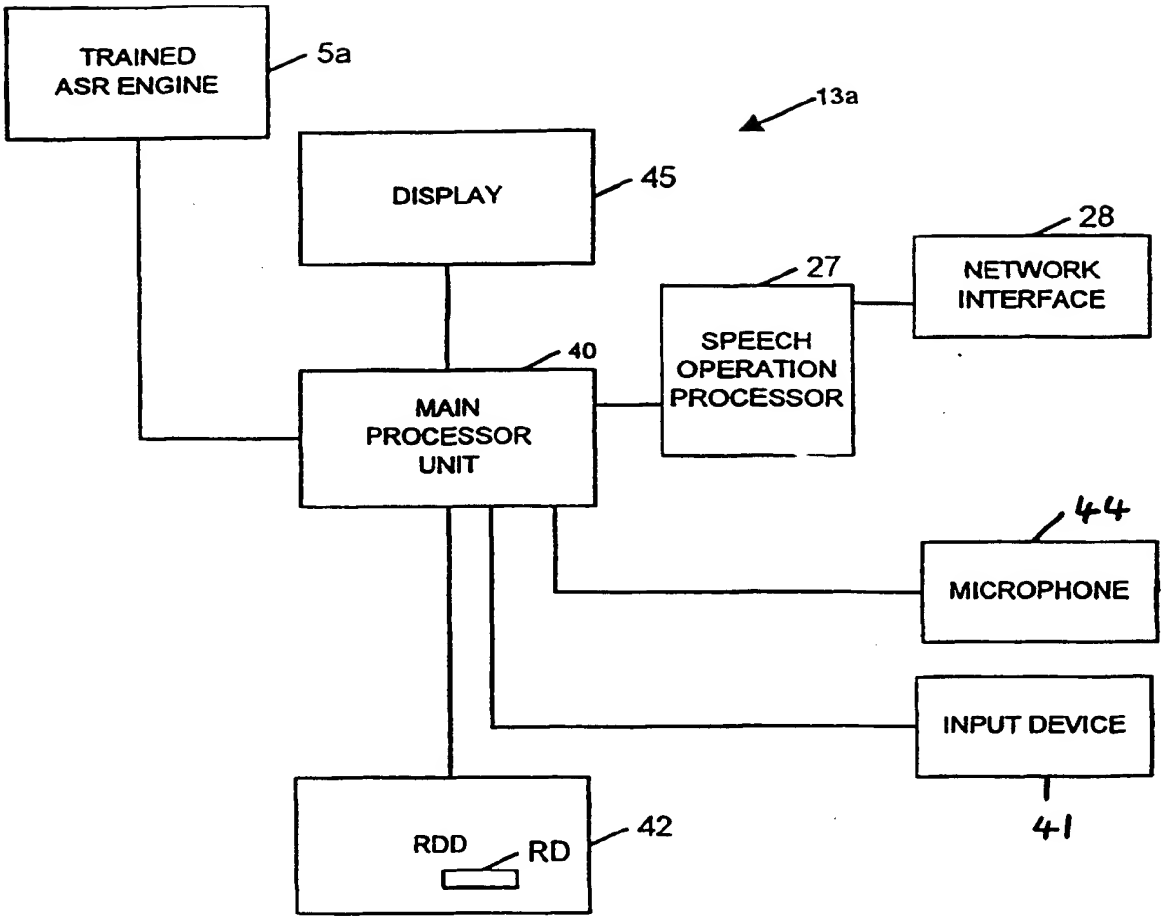

TRAINED ASR ENGINE
5a
13a
DISPLAY
45
28
NETWORK INTERFACE
27
SPEECH OPERATION PROCESSOR
40
MAIN PROCESSOR UNIT
44
MICROPHONE
INPUT DEVICE
41
RDD
RD
42

FIG. 27

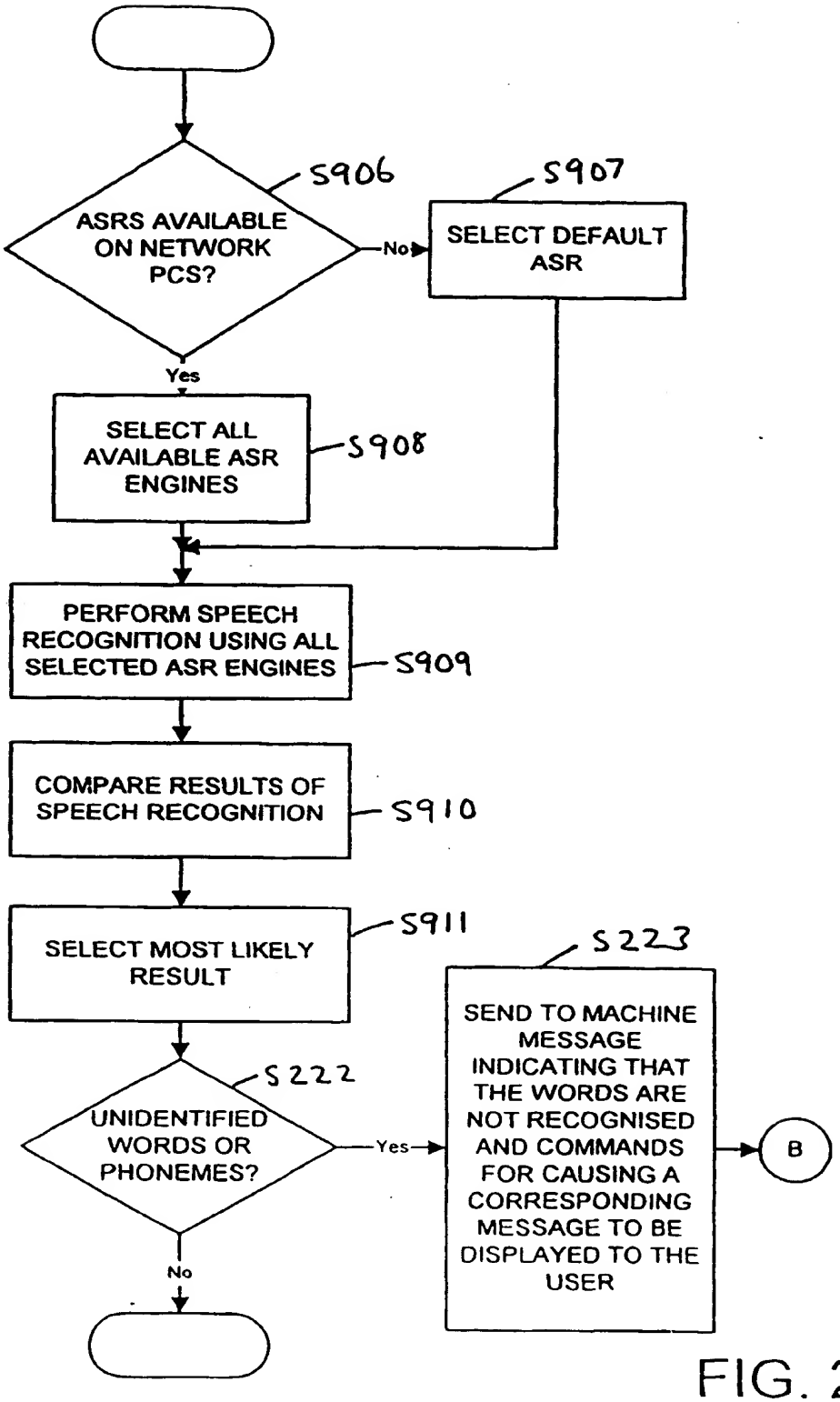

S906
ASRS AVAILABLE ON NETWORK PCS?
No
S907
SELECT DEFAULT ASR
Yes
SELECT ALL AVAILABLE ASR ENGINES
S908
PERFORM SPEECH RECOGNITION USING ALL SELECTED ASR ENGINES
S909
COMPARE RESULTS OF SPEECH RECOGNITION
S910
SELECT MOST LIKELY RESULT
S911
S222
UNIDENTIFIED WORDS OR PHONEMES?
Yes
S223
SEND TO MACHINE MESSAGE INDICATING THAT THE WORDS ARE NOT RECOGNISED AND COMMANDS FOR CAUSING A CORRESPONDING MESSAGE TO BE DISPLAYED TO THE USER
B
No

FIG. 28

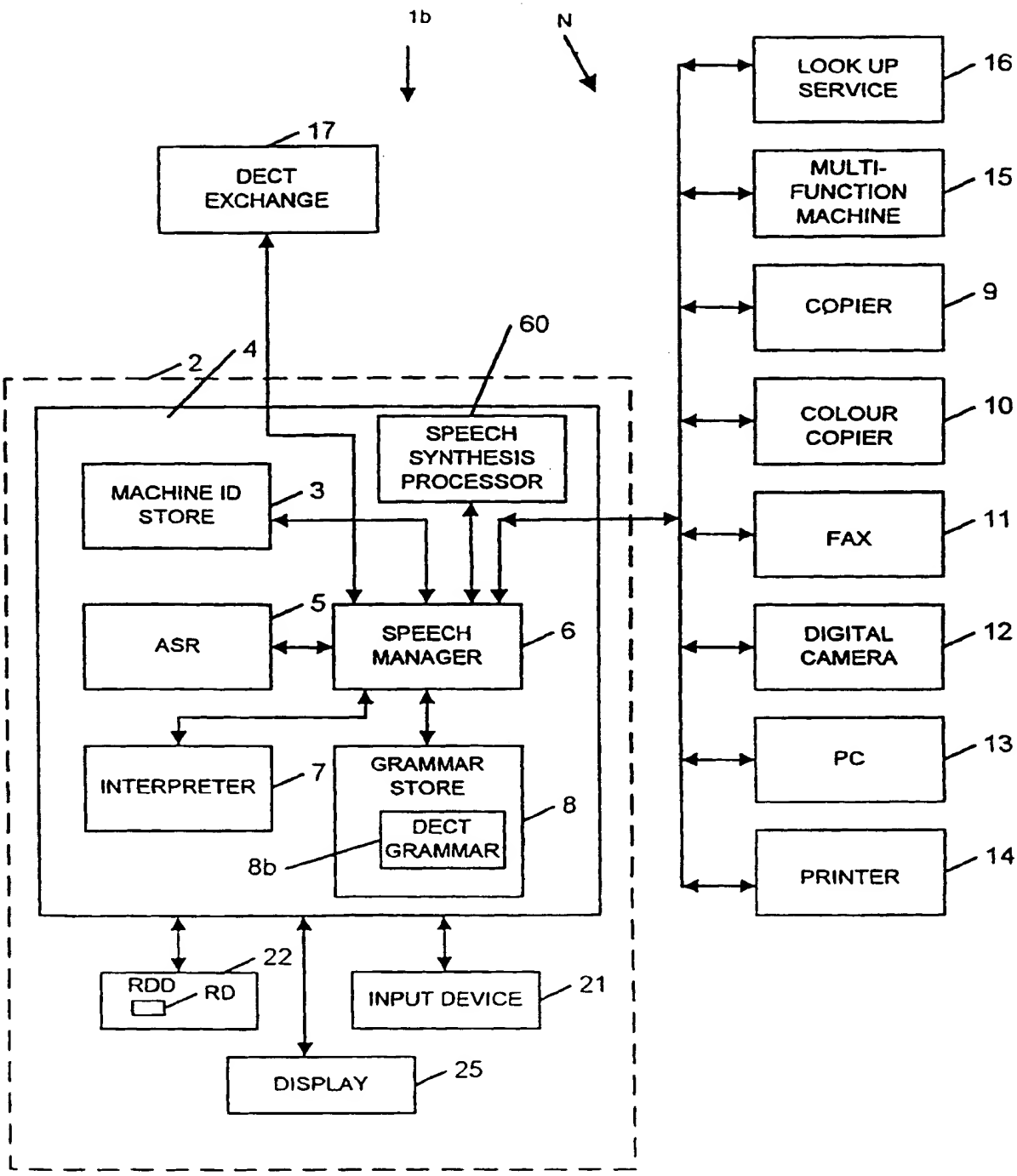
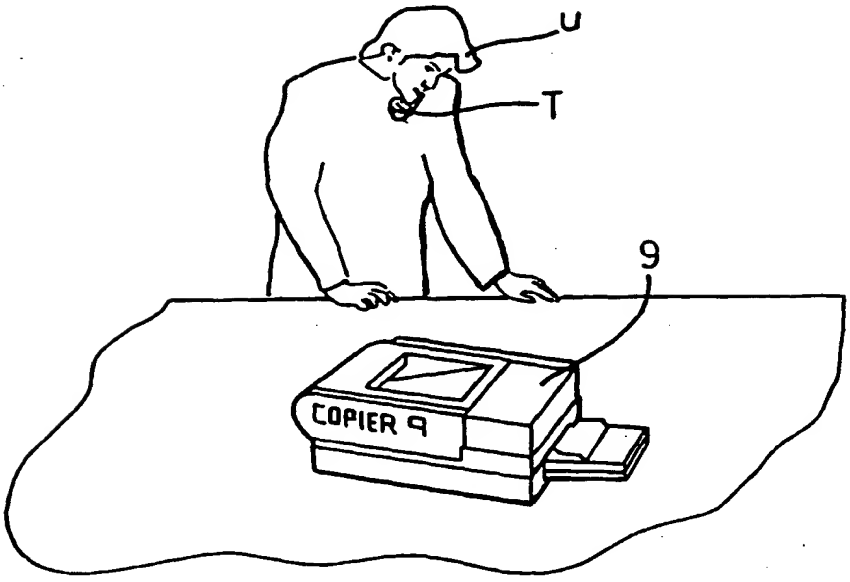

1b
N
LOOK UP SERVICE
16
DECT EXCHANGE
17
MULTI-FUNCTION MACHINE
15
COPIER
9
60
2
4
SPEECH SYNTHESIS PROCESSOR
COLOUR COPIER
10
MACHINE ID STORE
3
FAX
11
5
ASR
SPEECH MANAGER
6
DIGITAL CAMERA
12
INTERPRETER
7
GRAMMAR STORE
8
PC
13
DECT GRAMMAR
8b
PRINTER
14
22
RDD
RD
INPUT DEVICE
21
DISPLAY
25

FIG. 29

FIG. 30



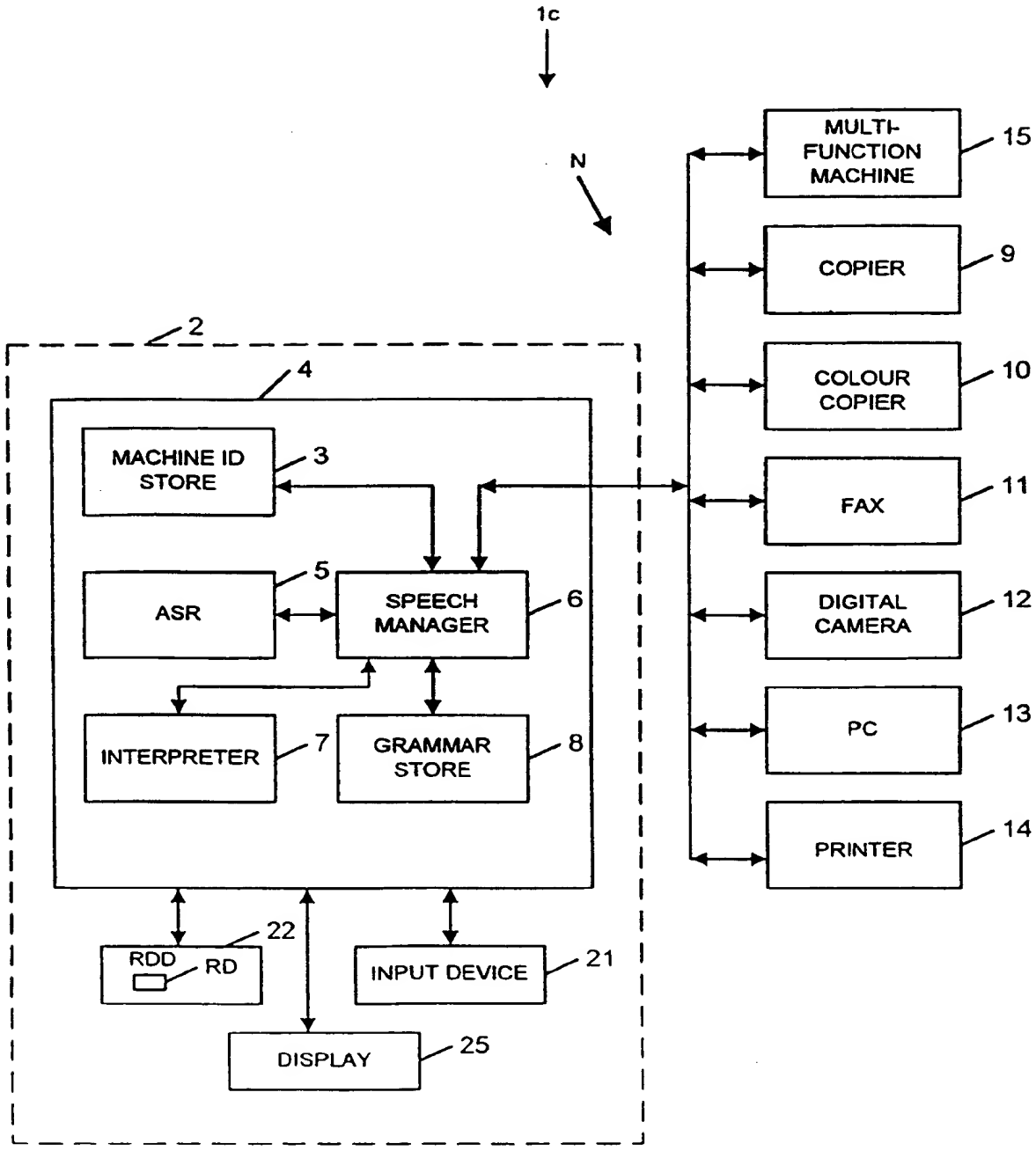

1c
N
MULTI-FUNCTION MACHINE
15
COPIER
9
COLOUR COPIER
10
FAX
11
DIGITAL CAMERA
12
PC
13
PRINTER
14
2
4
MACHINE ID STORE
3
ASR
5
SPEECH MANAGER
6
INTERPRETER
7
GRAMMAR STORE
8
RDD
RD
22
INPUT DEVICE
21
DISPLAY
25

FIG. 31

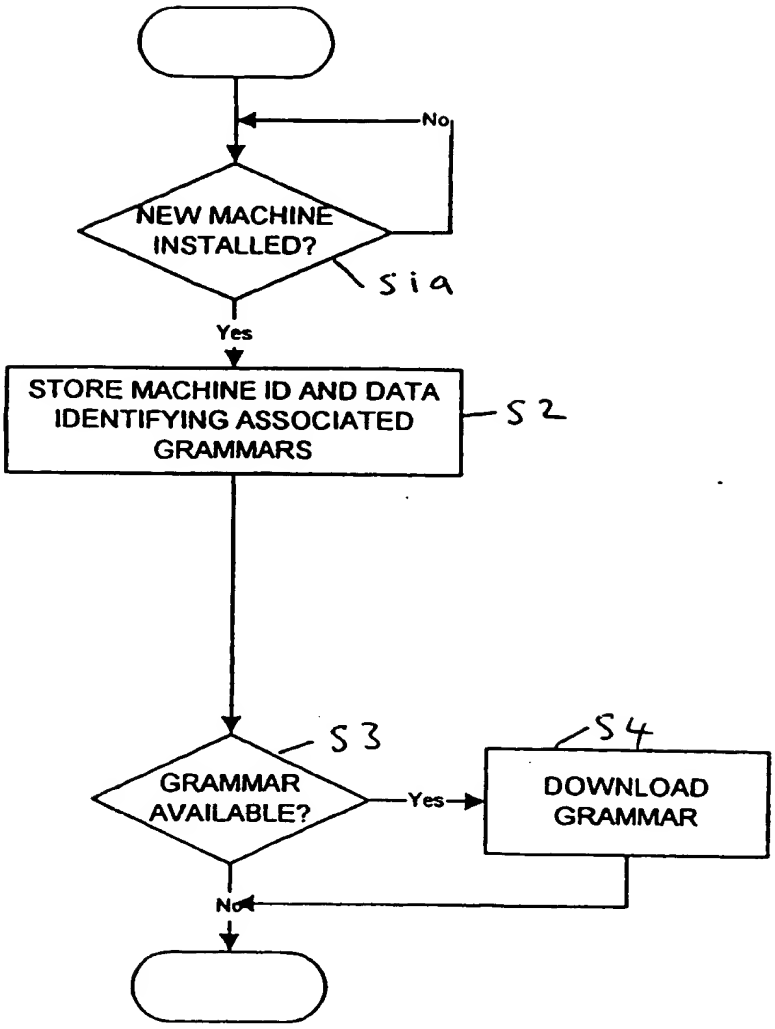

No
NEW MACHINE INSTALLED?
S1a
Yes
STORE MACHINE ID AND DATA IDENTIFYING ASSOCIATED GRAMMARS
S2
S3
GRAMMAR AVAILABLE?
Yes
S4
DOWNLOAD GRAMMAR
No

FIG. 32

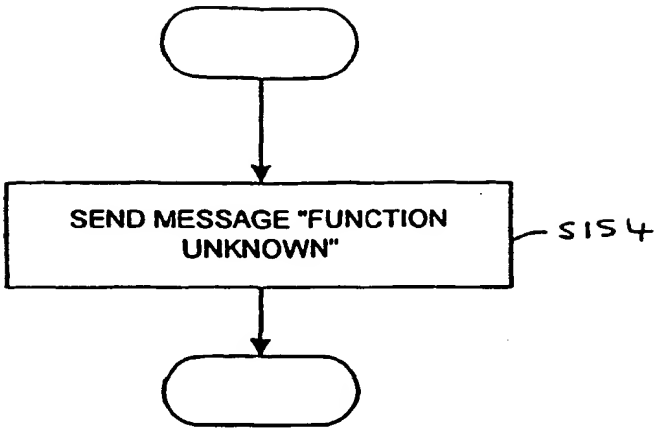

SEND MESSAGE "FUNCTION UNKNOWN"
S154

FIG. 33

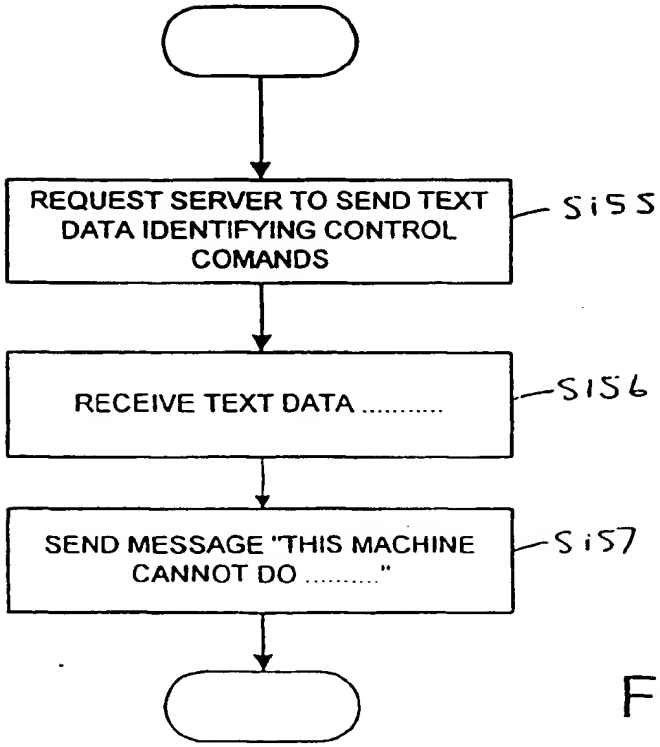

REQUEST SERVER TO SEND TEXT DATA IDENTIFYING CONTROL COMANDS
S155
RECEIVE TEXT DATA .....
S156
SEND MESSAGE "THIS MACHINE CANNOT DO ....."
S157

FIG. 34

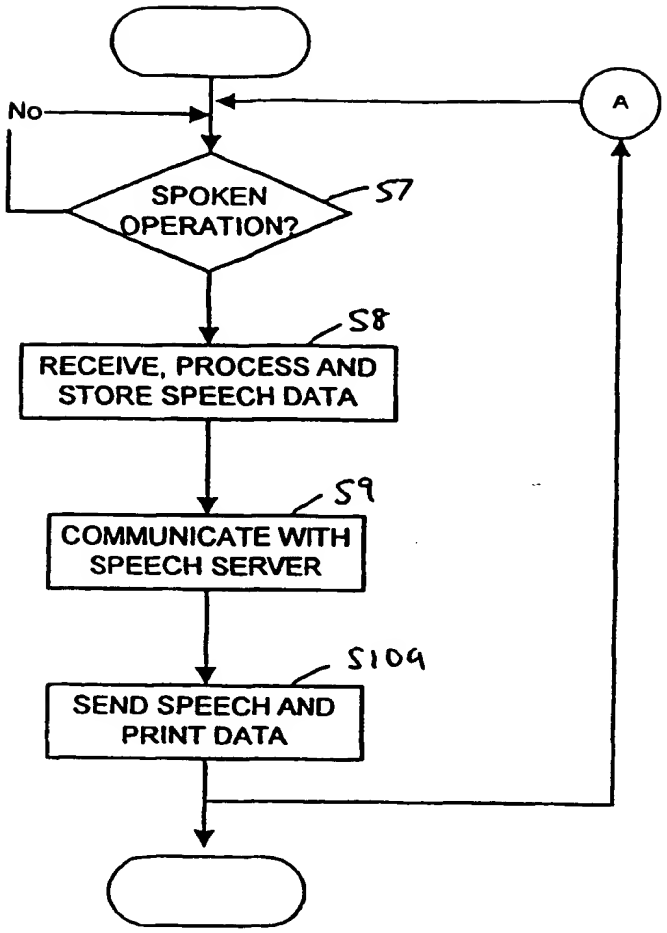

No
A
SPOKEN OPERATION?
S7
S8
RECEIVE, PROCESS AND STORE SPEECH DATA
S9
COMMUNICATE WITH SPEECH SERVER
S109
SEND SPEECH AND PRINT DATA

FIG. 35

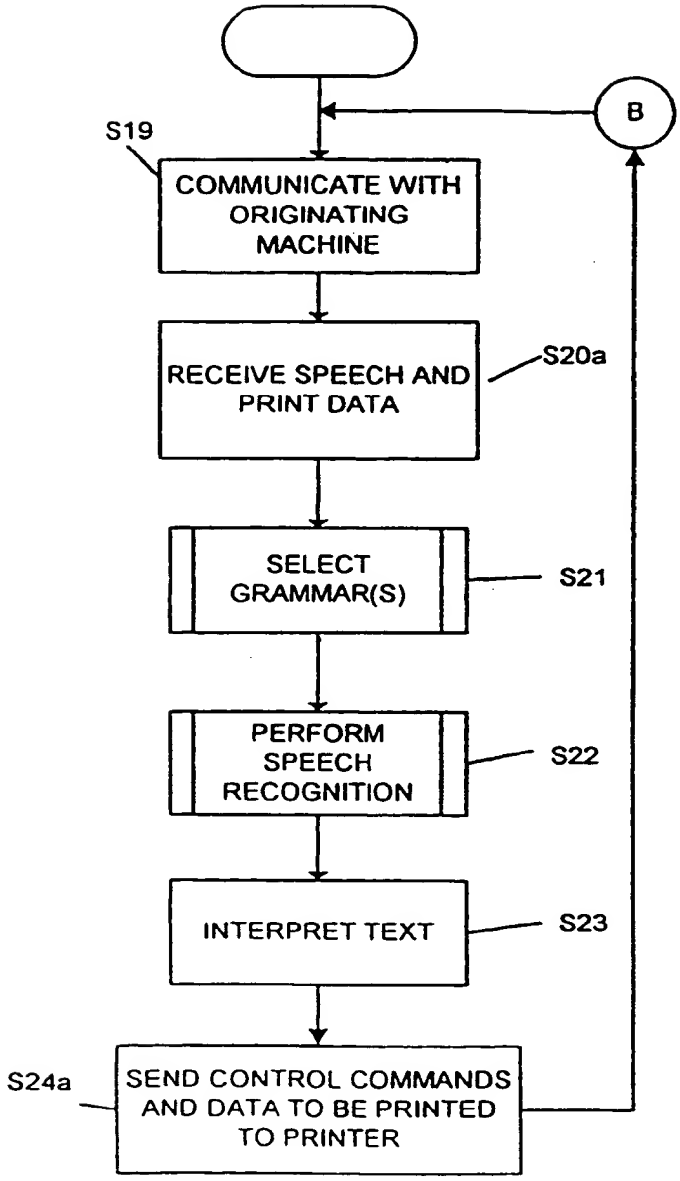

B
S19
COMMUNICATE WITH ORIGINATING MACHINE
RECEIVE SPEECH AND PRINT DATA
S20a
SELECT GRAMMAR(S)
S21
PERFORM SPEECH RECOGNITION
S22
INTERPRET TEXT
S23
S24a
SEND CONTROL COMMANDS AND DATA TO BE PRINTED TO PRINTER

FIG. 36